Performance Management

CHAPTER

5

CHAPTER OUTLINE

The Role of I/O Psychology in Performance Management
- Uses of Performance Appraisal
- Sources of Performance Ratings

Practitioner Forum **Laura Finfer**
- Rating Formats
- Rating Errors

I/O Today **Should We Eliminate Performance Ratings?**
- Rater Considerations

Contemporary Performance Appraisal Research
- The Importance of the Social-Psychological Context
- Providing Performance Feedback
- Legal Issues in Performance Appraisal

Summary

Photo: Chabruken/Getty Images

LEARNING OBJECTIVES

This chapter should help you understand:

- The importance of performance management
- The purposes of performance appraisal
- The various formats used in the evaluation of performance
- The effect of rating errors on the appraisal process
- How the broader context in which performance appraisal takes place has various and diverse implications for performance appraisal and organizational functioning
- The role of employee development in the performance management process
- The important role played by legal issues in the performance appraisal process
- The complexities involved in giving and receiving performance feedback

In the previous chapter, we discussed issues related to the development and implementation of criteria that are used to describe what is done on a particular job. In this chapter, we discuss how these criteria are used to appraise performance as part of a system to manage and improve performance. Performance management can be described as a "continuous process of identifying, measuring, and developing the performance of individuals and teams and aligning performance with the strategic goals of the organization" (Aguinis, 2013, p. 2). Having our performance managed, including the process of appraisal, is something we've all experienced at one time or another. Of course, we may not have fond memories of our fathers telling us that we didn't weed the vegetable garden well enough or of our Psychology 100 professor telling us that our test scores so far gave us only a C in the course. The types of performance appraisal just described are examples of negative appraisals. But sometimes we did a nice job mowing the lawn or writing a term paper—instances in which we probably received positive appraisals.

There are many domains in which we are held accountable for our performance, but in this chapter we will talk specifically about the role of performance management and appraisal in organizational life—in which it plays a very important role indeed. As you'll soon see, this is a crucial area for any organization, and a fertile area for I/O practitioners.

THE ROLE OF I/O PSYCHOLOGY IN PERFORMANCE MANAGEMENT

I/O psychologists play a significant role in the area of performance management. They are often hired to help develop and implement performance management systems. I/O psychologists have measurement expertise, as well as a background in both human resources and organizational psychology—areas of knowledge that are integral to successful performance management. Many companies have I/O psychologists in their HR departments who are responsible for **performance management**, which I formally define as a motivational system of individual performance improvement (DeNisi & Pritchard, 2006). This system typically includes (1) objective goal setting, (2) continuous **coaching** and feedback, (3) performance appraisal, and (4) developmental planning. The key points here are twofold: These four components are linked to the company's goals and objectives, and the system is implemented on a continuous cycle rather than just once per year. Some research has proposed that when an organization is profitable, it is typically more willing to reinvest in HR practices like performance management, which then impacts employees, supervisors, and the organization itself (den Hartog, Boselie, & Paauwe, 2004). Remember that performance appraisal stems directly from the job analysis (refer back to Figure 3.1). Performance criteria are identified by the job analysis and used as the central element of the performance appraisal system. Without a careful job analysis, we would likely end up with unimportant or job-irrelevant criteria and appraising performance on the wrong criterion dimensions.

performance management
A system of individual performance improvement that typically includes (1) objective goal setting, (2) continuous coaching and feedback, (3) performance appraisal, and (4) development planning.

coaching
One-on-one collaborative relationship in which an individual provides performance-related guidance to an employee.

performance appraisal
Systematic review and evaluation of job performance.

Uses of Performance Appraisal

Performance appraisal, defined as the systematic review and evaluation of job performance, as well as the provision of performance feedback, is a key component of performance management and one of the most important processes conducted in organizations. It has many purposes, of which the three most significant are discussed here. First, performance appraisals are used to make important *personnel decisions,* such as who gets promoted, fired, demoted, or laid off; who gets a large raise, a small raise, or no raise at all; and so on. In efficient organizations, these decisions are not made haphazardly; they are made on the basis of performance appraisal data.

Second, performance appraisals are used for *developmental purposes.* Employees are informed of their performance strengths and weaknesses so that they can be proud of what they are doing well and can focus their efforts on the areas that need some work. On the whole, organizations benefit when employees perform better, and performance appraisal data are used to help employees become better performers. In addition, organizations are interested in seeing their employees advance within the company to other important jobs—an outcome that performance appraisal can facilitate. For example, an employee may be told that she needs to improve her interpersonal skills so that she will be eligible when the next promotion becomes available.

A third purpose of performance appraisal is what I'll call *documentation* of organizational decisions—a purpose that has recently evolved out of personnel decisions and the growing area of personnel law. Now that companies are very aware of the possibility of being sued over personnel business decisions, managers are increasingly using performance appraisals to document employees' performance patterns over time. In cases in which employees are fired for inadequate performance, the organization—if it has kept careful track—can point to detailed accounts of the employees' inferior performance, making it difficult for them to claim that they were fired without

EMPLOYEE
OF
THE MONTH

"It's one of the advantages of having your own business."

Thomas Bros./Cartoonstock.com

A CLOSER LOOK

Performance can be measured in many different ways. How has your performance been measured or recognized in any jobs that you've had, and what made that process useful for you?

just cause. (We will discuss legal issues in performance appraisal later in this chapter, and personnel law in general in Chapter 7.)

On the other hand, performance appraisals that are not carefully developed and implemented can have negative repercussions for both the organization and its employees. For instance, a poorly conceived appraisal system could get the wrong person promoted, transferred, or fired. It could cause feelings of inequity on the part of good employees who erroneously receive smaller raises than bad employees. It could lead to lawsuits in which the company has a very weak defense for why a particular individual was not promoted. Also, it could result in disgruntled employees who decrease their effort, ignore the feedback, or look for other jobs. Even customers are poorly served when an ineffective appraisal system causes employees to operate at less than their peak level of efficiency. Indeed, an ineffective performance appraisal system has widespread implications for everyone involved with the organization, which is why performance appraisal has received so much research attention (for reviews, see Levy & Williams, 2004; Schleicher et al., 2018; Schleicher, Baumann, Sullivan, & Yim, 2019). A good performance appraisal system is well received by ratees, is based on carefully documented behaviors, is focused on important performance criteria, is inclusive of many perspectives, and is forward looking with a focus on improvement.

Researchers who specialize in performance appraisal pursue research questions such as the following: (1) What is the best format or rating scale for performance appraisals? (2) To what extent do rater errors and biases affect the appraisal process? (3) How should raters be trained so that they can avoid these errors and biases? (4) What major contextual variables affect the appraisal process? (5) How important is the organizational context or culture in the appraisal process? (6) What factors affect how ratees and raters react to performance appraisal? In addition to addressing such questions, I/O psychology, as an empirically based applied discipline, attempts to use basic psychological principles to help organizations develop and implement motivating, fair, accurate, and user-friendly appraisal systems.

Sources of Performance Ratings

Performance feedback can be generated and delivered by various sources. Traditionally, supervisors were charged with conducting the performance appraisal and delivering the performance feedback. This top-down approach continues to be very popular, and it is quite common for an organization to include this type of appraisal as part of the performance management process. However, there are other, more contemporary approaches for the use of feedback sources, with multisource feedback being the most prevalent.

Multisource Feedback This method, sometimes called **360-degree feedback** (Fletcher & Baldry, 2015), involves multiple raters at various levels of the organization who evaluate and provide feedback to a target employee. As presented in **Figure 5.1,** these multiple sources typically include subordinates (or direct reports), peers, supervisors, customers, and clients and even self-ratings. Note the variation that exists between the different raters for each dimension.

360-degree feedback
A method of performance appraisal in which multiple raters at various levels of the organization evaluate a target employee and the employee is provided with feedback from these multiple sources.

These systems have become increasingly important to the modern organization in terms of performance assessment and management (Hoffman, Lance, Bynum, & Gentry, 2010). Many companies—such as Home Depot, Procter & Gamble, General Electric, Intel, and Boeing, among others—have used them for a variety of purposes that are consistent with both greater employee expectations and the more sophisticated organizations of the 21st century.

Three basic assumptions are held by advocates of 360-degree feedback systems. First, when multiple raters are used, the participants are happier because they are involved in the process—and this calls to mind the importance of participation alluded to earlier. Second, and perhaps more important, when multiple raters from different levels of the organization rate the same target employee, the idiosyncrasies (and biases) of any single rater are overcome. For instance, if my supervisor doesn't like me and rates me severely for that reason, additional ratings from other individuals that aren't severe should overcome my supervisor's rating and highlight the possibility that there may be a problem with that rating. Third, multiple raters bring with them multiple

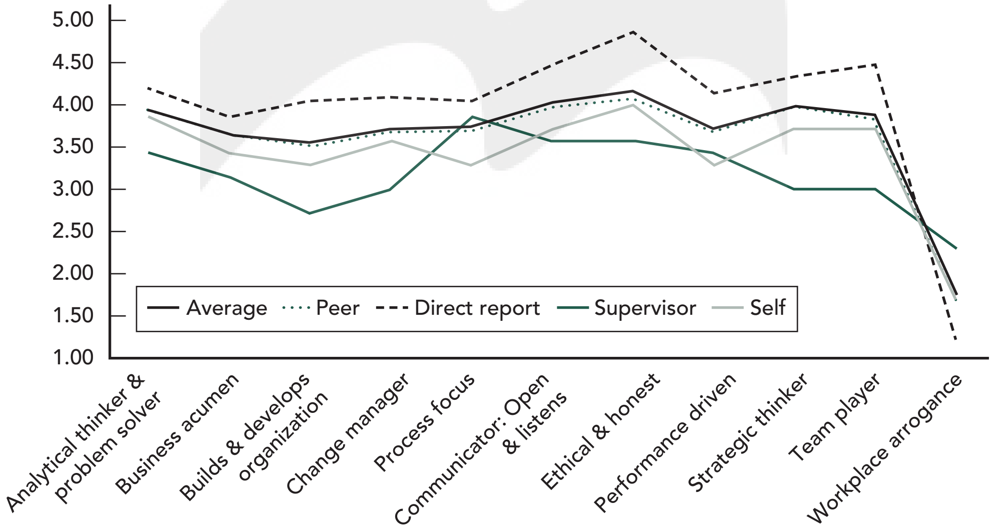

FIGURE 5.1 360-Degree Development Report This sample figure from a 360-degree report shows the scores provided by four raters (and one aggregate score) across 11 "leadership" dimensions on a 5-point scale.

Source: Human Resource Decisions, Inc.

upward appraisal ratings
Ratings provided by individuals whose status, in an organizational-hierarchy sense, is below that of the ratees.

perspectives on the target employee, allowing for a broader and more accurate view of performance. For instance, universities often require that students rate faculty teaching because it is believed that they have a valuable perspective to share about teaching effectiveness. These are sometimes called **upward appraisal ratings** because they refer to ratings provided by individuals whose status is, in an organizational-hierarchy sense, below that of the ratee (Atwater, Daldman, Atwater, & Cartier, 2000). A recent study that included traditional supervisor ratings, peer ratings, and subordinate (upward) ratings demonstrated that these perspectives differ from each other regarding managerial competencies and how they predict managerial effectiveness (Semeijn, Van der Heijden, & Van der Lee, 2014). The authors concluded that the multisource approach is beneficial in assessing both competencies and effectiveness.

Although 360-degree feedback has been used in organizations for a few decades, research efforts to understand it have only recently begun to catch up with practitioner usage. Until recent years the majority of empirical work on 360-degree feedback has focused on measurement properties of the ratings, such as the extent of agreement among rating sources (Levy & Williams, 2004). However, we are starting to see considerably more research on 360-degree feedback on broader contextual issues. In addition, there has also been an influx of research focused on applications to the health field (doctors, nurses, medical residents) and education (teachers, principals); for a review, see Stevens, Read, Baines, Chatterjee, and Archer (2018). A recent study examined the use of 360-degree feedback with 385 surgeons (Nurudeen et al., 2015). Results included very high percentages of surgeons reporting that the feedback they received was accurate and reporting that they made changes to their practice as a result of the feedback. Similarly, a large percentage of department heads reported that they believed the feedback provided to their surgeons was accurate. Almost 75% of the participants (raters and ratees) found the process valuable, and over 80% were willing to participate in future 360-degree evaluations. There is even an app (called Healthcare Supervision Logbook) that allows doctors who are training medical students to provide feedback to the students following a clinical session. The trainees can use the app to provide feedback about the curriculum and training to the doctor trainers (Gray, Hood, & Farrell, 2015). It can also be used to gather patient and peer feedback.

In another recent investigation, this time in the education arena, principals were evaluated with a 360-degree instrument and researchers studied how they reacted to conflicting feedback among sources (Goldring, Mavrogordato, & Haynes, 2015). They found that principals experienced cognitive dissonance (i.e., discomfort and tension) when the teacher ratings of their performance were lower than their own ratings of their performance. The authors assert that principals are motivated to reduce the dissonance and that they can do that by either working harder and improving how they do their jobs or by discounting the teacher ratings. The researchers suggest that principals need to be trained in how to receive, evaluate, and use the feedback they are given.

One firm, Human Resource Decision, Inc., employs a 360-degree feedback system that consists of four main tools: (1) a 360-degree development questionnaire, which is administered to multiple rating sources and measures 13 skill dimensions, such as leadership and business acumen; (2) a 360-degree feedback report, which provides the results of the ratings from the various sources, as well as some summary information about ratees' strengths and weaknesses; (3) a development workbook, which helps employees work with and understand the feedback report; and (4) a development guide, which provides suggested readings and activities to improve skills in the targeted areas. This 360-degree system has been used by many companies across a variety of industries.

PRACTITIONER FORUM

Courtesy of Laura Finfer

Laura Finfer

PhD, 1989, I/O Psychology,
Purdue University
Principal
Leadership Excellence Consulting

Most performance appraisal research (and practice) focuses below the C-suite—that is, it covers employees below the level of the CEO and the CEO's direct reports. And with good reason: The bulk of the employee population does not reside in the C-suite. Performance of the C-suite is traditionally reflected in business metrics that are clear and often publicly communicated (e.g., revenue, profit, market share, etc.). Yet, the higher up one is in an organization, the greater the feedback vacuum. One way this is addressed is via a Verbal 360 process.

Verbal 360s (also called Live 360 or Stakeholder Reviews) involve interviewing those around the executive, including direct reports, peers, internal stakeholders, and the CEO. Interviews are conducted by skilled practitioners who customize the questions to the specific business context and to the executive's role in achieving desired business outcomes. This format allows interviewers to confidentially gather critical incidents and categorize the information thematically (e.g., strategic thinking, innovation, results orientation). For example, I commonly hear in interviews that an executive is "smart." Probing for an example of when that executive demonstrated "smarts" provides rich data from which other insights can be gleaned (e.g., the executive may explore a range of solutions, connect data in counterintuitive ways, or dig deeply for relevant information).

The interviewer is also in a unique position to objectively synthesize the information gathered. On one occasion, I interviewed stakeholders about the CFO of an organization. Comments frequently linked this CFO to the CEO, who had worked with the CFO in several firms. Observations of this link were so prevalent that I concluded that the CFO operated in the shadow of the CEO; the feedback discussion highlighted the need for the CFO to establish his own identity and put a unique handprint on the firm. By nature of the specificity of the feedback, the executive pinpoints what to continue and what to do differently.

The feedback and analysis from Verbal 360s can have a profound impact. I worked with an organization where the CEO requested that his team participate in a Verbal 360 process. The team was relatively new (in place for three years) and confronting significant industry-wide economic challenges. Upon receipt of the feedback, each executive discussed his or her role in achieving the CEO's vision. In addition, aggregating themes across executives provided the CEO with a snapshot of the culture and insights about his own leadership. Based on this data, I advised the CEO to leverage the team to participate in strategic planning and brainstorm new ideas—both critical activities for enrichment in a challenging business environment.

Applying Your Knowledge

1. At any level, the idiosyncrasies of raters must be accounted for when conducting 360-degree feedback. Considering their high profiles within an organization, what unique issues might C-suite executives face in this respect?
2. Imagine you are asked to construct a 360-degree feedback system for the CEO of a company. What steps could you take to help ensure that the results and analysis of that feedback are effectively used by the company?

What does the future hold for 360-degree feedback? Consistent with this chapter's theme of focusing on the social context of appraisal is a list of recommendations to follow for implementing 360-degree feedback: This list includes (1) being honest about how the ratings will be used, (2) helping employees interpret and deal with the ratings, and (3) avoiding the presentation of too much information (DeNisi & Kluger, 2000). The frequency with which 360-degree feedback is used in organizations suggests that it will continue to play an important role. Recent research suggests that the following important issues will continue to attract attention: construct validity of ratings, determinants of multisource ratings (such as rating purpose, liking, and personality), and the effect of multisource ratings on employee attitudes and development.

telework
Working arrangements in which employees enjoy flexibility in work hours and/or location.

Challenges in Telework Another growing trend in organizations is the increased frequency of **telework**, employees working from home or some other remote location, a practice that is growing rapidly (Golden, Barnes-Farrell, & Mascharka, 2009). Telecommuting alters communication. Many employee–employee and employee–supervisor interactions don't take place face-to-face but via e-mails, phone calls, messaging apps, conference calls, and a host of online communication platforms, including Google Hangouts, Slack, Microsoft Outlook, or a company's own internal messaging system. This arrangement suggests that supervisors doing performance appraisal must rely on indirect sources of performance information, like gathering information from those who work directly with the employee and reviewing written documentation of work, instead of direct interactions and on-the-job observations. Recent research has demonstrated that when provided with both direct and indirect performance information, supervisors rely more on direct performance information (Golden et al., 2009). Supervisors' tendencies to downplay the indirect information that is frequent in telework suggest potential room for performance appraisal errors and ineffectiveness. (See Chapter 11 for information on telework and worker well-being.)

Of course, one important aspect of telecommuting is its relationship with performance. Although there hasn't been a lot of empirical evidence in this area, there is some recent research. For instance, a survey of 273 supervisor–subordinate dyads found that telecommuting did have a positive effect on performance. The strength of this effect was moderated by various factors, but it's important to note that the effect was never negative (Golden & Gajendran, 2014). That is to say, telecommuting may have a weak or strong positive effect on performance depending on various situational characteristics, but it never has a negative effect. As more evidence of the positive effect of telecommuting on performance, Gajendran, Harrison, and Delaney-Klinger (2015) studied 323 supervisor–subordinate dyads and found that telecommuting was positively associated with both task performance and contextual

Jetta Productions/Getty Images

A CLOSER LOOK
What factors might play a role in determining whether telecommuting has a positive or negative effect on performance?

performance. Further, the effect of telecommuting on performance was strengthened when there was a positive relationship between the subordinate and supervisor. When an employee who experiences a favorable subordinate–supervisor relationship is given the freedom to telecommute, that employee will assume more autonomy, which will be reflected in higher levels of task and contextual performance. Indeed, a recent study of employees in Great Britain was able to demonstrate that telework, home-based work, and flexible time caused the firm's overall financial performance and labor productivity to improve (Giovanis, 2018).

Rating Formats

When it comes time for an evaluator to appraise someone's performance, he or she typically uses some type of rating form. There are quite a few options here; in this section, we will discuss the ones most frequently employed.

Graphic Rating Scales Graphic rating scales are among the oldest formats used in the evaluation of performance. These scales consist of a number of traits or behaviors (e.g., dependability), and the rater is asked to judge how much of each particular trait the ratee possesses or where on this dimension the ratee falls with respect to organizational expectations. Today, graphic rating scales usually include numerical/verbal anchors at various points along the scale, such as "1/below expectations," "4/meets expectations," and "7/exceeds expectations," and the score is whatever number is circled. Graphic rating scales are commonly used in organizations due, in part, to the ease with which they can be developed and used. **Figure 5.2** provides an example of a graphic rating scale—in this case, one that appraises the extent to which employees are "following procedures."

Behaviorally Anchored Rating Scales Behaviorally anchored rating scales, or BARS (Smith & Kendall, 1963), are similar to graphic rating scales except that they provide actual behavioral descriptions as anchors along the scale. An example of a 9-point BARS for a nuclear power plant operator's tendency to "follow procedures" is shown in **Figure 5.3.**

BARS

A performance appraisal format that uses behavioral descriptors for evaluation.

BARS are perhaps best known for the painstaking process involved in their development (see Smith & Kendall, 1963). This process can be summarized as taking place in five steps (Bernardin & Beatty, 1984). First, a group of participants—such as employees, supervisors, or subject matter experts (SMEs)—identifies and carefully defines several dimensions as important for the job ("follows procedures," "works in a timely manner," etc.). Second, another group of participants generates a series

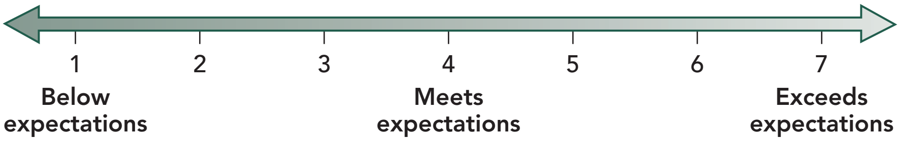

FIGURE 5.2 Graphing Rating Scale for "Following Procedures"

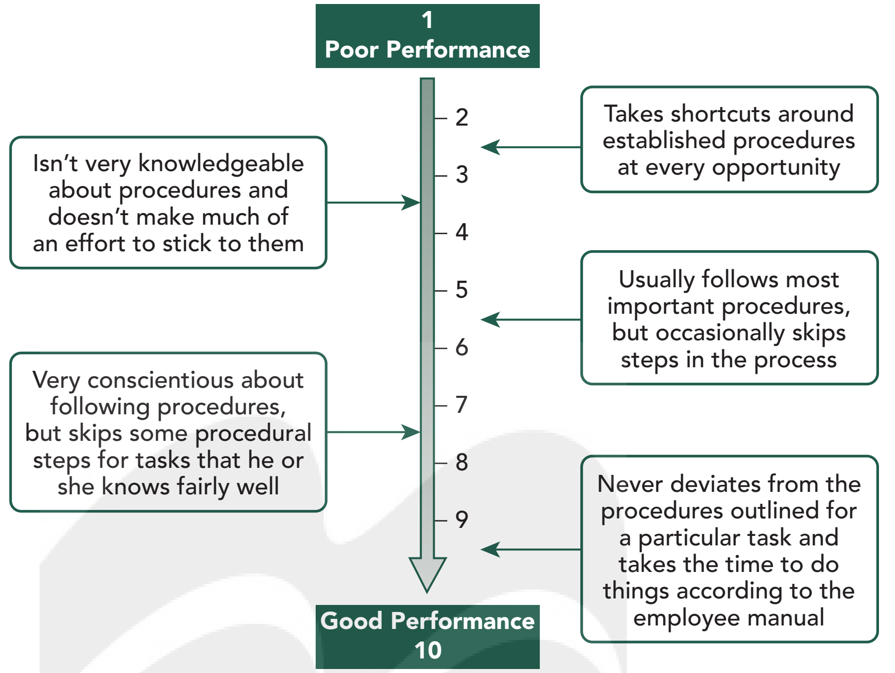
1
Poor Performance
2
3
4
5
6
7
8
9
Good Performance
10
Takes shortcuts around established procedures at every opportunity
Isn't very knowledgeable about procedures and doesn't make much of an effort to stick to them
Usually follows most important procedures, but occasionally skips steps in the process
Very conscientious about following procedures, but skips some procedural steps for tasks that he or she knows fairly well
Never deviates from the procedures outlined for a particular task and takes the time to do things according to the employee manual

FIGURE 5.3 Behaviorally Anchored Rating Scale for "Following Procedures"

of behavioral examples of job performance (similar to the items in Figure 5.3) for each dimension. These behavioral examples are called **critical incidents**. Participants are encouraged to write critical incidents at high-, medium-, and low-effectiveness levels for each dimension. In Figure 5.3, for example, "Never deviates from the procedures outlined for a particular task and takes the time to do things according to the employee manual" represents high effectiveness, whereas "Takes shortcuts around established procedures at every opportunity" represents low effectiveness.

critical incidents
Examples of job performance used in behaviorally anchored rating scales or job-analytic approaches.

Third, yet another group of participants is asked to sort these critical incidents into the appropriate dimensions. During this *retranslation* stage, the goal is to make sure that the examples generated for each dimension are unambiguously associated with that dimension. Usually a criterion, such as 80%, is used to weed out items that are not clearly related to a particular dimension: If fewer than 80% of the participants place the critical incident in the correct dimension, the item is dropped. A fourth group of participants then rates each remaining behavioral example on its effectiveness for the associated dimension. This rating is usually done on a 5- or 7-point scale. Any item with a large standard deviation is eliminated because a large standard deviation indicates that some respondents think the item represents effective performance, whereas others think it represents ineffective performance—obviously, a problematic situation. Fifth, items that specifically represent performance levels on each dimension are chosen from the acceptable pool of items, and BARS are developed and administered.

This very deliberate and thorough development process is both the biggest strength and the biggest weakness of BARS. Usually such a detailed process involving at least four different groups of participants results in a useful and relevant scale. However, the costs in both time and money are too high for many organizations, so they often employ a simple graphic scale instead.

Checklists Checklists are another popular format for performance appraisals. Here, raters are asked to read a large number of behavioral statements and to check off each behavior that the employee exhibits. One example from this category is the *weighted checklist,* which includes a series of items that have previously been weighted as to importance or effectiveness; specifically, some items are indicative of desirable behavior, whereas others are indicative of undesirable behavior. **Figure 5.4** presents an example of a weighted checklist for the evaluation of a computer and information systems manager.

In a real-life situation, the scale shown in Figure 5.4 would be modified in two ways. First, the items would be scrambled so as not to be in numerical order; second, the scale score column would not be part of the form. To administer this type of scale, raters would simply check the items that apply; in our example, the sum of items checked would be the computer and information systems manager's performance appraisal score. Note that as more and more of the negative items are checked, the employee's summed rating gets lower and lower.

Forced-choice checklists are also used by organizations, though not as frequently as weighted checklists. Here, raters are asked to choose two items from a group of four that best describe the target employee. All four appear on the surface to be favorable, but the items have been developed and validated such that only *two* are actually

Computer and Information Systems Manager

Check Box (if applicable)	Behavioral Descriptor	Scale Score
☐	Effectively manages backup and computer help systems	+8.5
☐	Consults with users, vendors, and technicians on a regular basis	+6.7
☐	Stays abreast of technology advancements	+4.4
☐	Maintains 40 hours per week in the office	+1.5
☐	Doesn't anticipate likely problems	–1.2
☐	Pays little attention to financial details	–2.4
☐	Rarely provides leadership to his/her work group	–4.6
☐	Isn't able to implement necessary changes	–6.7
	TOTAL SCORE:	

FIGURE 5.4 A Weighted Checklist

good discriminators between effective and ineffective performers. The purpose of this approach is to reduce purposeful bias or distortion on the part of raters. In other words, because all the items appear favorable and the raters don't know which two are truly indicative of good performance, they cannot intentionally give someone high or low ratings.

One drawback to this approach is that some raters don't like it because they feel as though they've lost control over the rating process. How would you feel if you had to choose two statements that describe your poorly performing subordinate, but all four statements seem positive? Researchers have recently developed an appraisal format that appears to reduce the bias sometimes associated with other approaches, but without the negative reactions on the part of the raters. Borman and his colleagues developed the computerized adaptive rating scale (CARS), which, although still relatively new to the performance appraisal field, seems to be more sound from a measurement perspective than are other approaches; it also provides more discriminability—that is, the scale does a better job of differentiating between effective and ineffective performers (Borman et al., 2001; Schneider, Goff, Anderson, & Borman, 2003). CARS is an ideal point response method (Drasgow, Chernyshenko, & Stark, 2010) in which raters are given two statements about performance with one slightly above average and one slightly below average. The rater is asked to choose the one that best reflects the ratee's performance. This is followed by two more statements—one more favorable and one less favorable than the point just chosen—and the rater chooses the best option. This continues until all the relevant items have been presented and the best performance estimate/rating has been determined (Borman, 2010).

Employee Comparison Procedures The final category of rating formats, *employee comparison procedures,* involves evaluation of ratees with respect to how they measure up to or compare with other employees. One example of this type of format is *rank-ordering,* whereby several employees are ranked from best to worst. Rank-ordering can be particularly useful for making promotion decisions and discriminating the very best employee from the rest. A second example, *paired comparisons,* involves the comparison of each employee with every other employee. If a manager has only three employees to evaluate, this isn't too difficult a task. (Think about doing this for each of your instructors this semester—comparing each to every one of the others.) However, as the number of ratees increases, so does the complexity of the task. Although this method is one way to arrive at a "clear winner" among employees, it obviously becomes very cumbersome as the number of employees grows.

A third example is called *forced distribution.* Here, raters are instructed to "force" a designated proportion of ratees into each of five to seven categories. This procedure is similar to "grading on the normal curve," whereby teachers assign grades based on meeting the normal curve percentages (i.e., 68% of the grades assigned are Cs, 13.5% are Bs, 13.5% are Ds, 2.5% are As, and 2.5% are Fs). Sometimes organizations require supervisors to use the same sort of procedure, resulting in

TECHNICAL TIP

The formula N(N – 1)/2 can be used to calculate the total number of comparisons. For example, whereas 3 employees would involve 3 comparisons, 10 employees would involve 45!

the categorization of one-third of the subordinates as average, one-third as below average, and one-third as above average (see the *Time* article "Rank and Fire" by Greenwald, 2001). This is often done because performance ratings are tied to raises and, with a limited pool of money for raises, the company wants to make sure that not too many employees are rated as eligible for a raise and to potentially remove inferior employees from the organization.

To appreciate how employees might feel about this approach, think about how you would feel if your psychology instructor told you that you were to be graded on a normal curve. In other words, if you had an average of 95 in the course but 3% of your classmates had an average of 96 or better, you would *not* receive an A because only 3% can get As. Needless to say, forced distribution is not a popular approach among ratees, whether students or employees, but it was very popular among Fortune 500 companies in the 1990s and early to mid-2000s. It's estimated that 20% of these organizations employed this practice in the mid-2000s (Grote, 2005), with companies such as General Electric, 3M, Texas Instruments, Microsoft, Ford, Goodyear, and Hewlett-Packard among the most ardent supporters. However, some very public lawsuits over the use of these systems have created a controversy regarding the extent to which underrepresented groups tend to be disproportionately ranked in the low category, resulting in adverse impact against these groups (Giumetti, Schroeder, & Switzer, 2015). *Adverse impact* is an important concept in personnel law and may indicate illegal discrimination against a particular group (see Chapter 7 for a more detailed discussion of this issue). Many thoughtful papers and analyses have pointed out other weaknesses in the forced distribution approach, which include the negative reactions of both raters and ratees, the fallacy of forcing some employees into the bottom rung regardless of their true performance, and problems with continuity, where, for example, an employee that was rated average or above average will eventually fall to the bottom rung as the workforce is strengthened and perceptions of effective performance change. For these and other reasons, we have recently seen many more companies back away from this approach. You can continue to follow these issues in the popular press, such as in Gaurav Gupta's 2018 piece on the *Forbes* website called "Are You Still Using Force Rankings? Please Stop."

Contemporary Trends in Rating Formats Although there has been less research and less written about rating formats in recent years, there are a couple of interesting contemporary trends. First, while there are many rating formats and most are quantitative in nature (i.e., performance is rated in terms of numbers), most of these formats include narrative comments (e.g., your performance on this dimension is above average, but you need to enhance your administrative skills) in some way. It is only recently that researchers have begun examining these narrative comments. Some work has found that supervisors' and subordinates' comments are considerably more specific than the comments from peers (Gillespie, Rose, & Robinson, 2006). A recent conceptual paper proposes a framework for the use of narrative comments in the performance appraisal process (Brutus, 2010) and identifies important characteristics, such as the specificity or breadth of the comments as well as the processes that raters and ratees employ in the use of these narrative comments. An even more recent study that used text mining (computer scoring) to derive meaning from narrative comments found that the narrative comments explained unique variance in performance ratings (Speer, 2018). These narrative comments also predicted the likelihood that employees would turn over in the next year.

Finally, additional outside-the-box thinking has resulted in the use of feedforward interviews (FFIs) for performance appraisal (Kluger & Nir, 2010). This idea is borrowed from appreciative inquiry (see Chapter 14), a part of the positive psychology movement. The FFI is proposed to replace the traditional performance appraisal interview and to facilitate positive change by focusing on employees' strengths, rather than weaknesses, and also to enhance the relationship between the rater and ratee. While promising, there is very little empirical research on the effectiveness of the FFI. However, one new study trained customer service managers in an FFI approach that focused on employees' attention to positive work experiences in which goals were met and success was attained. Researchers found that the performance of subordinates whose managers employed FFI improved more over a four-month period than did those whose managers used more traditional performance appraisal techniques with their employees (Budworth, Latham, & Manroop, 2015). Obviously, more research is needed, but there is some potential for organizations to consider feedforward as an alternative or in combination with feedback.

An Evaluation of the Various Alternative Methods In more recent years, research has looked at how personality and format may impact performance ratings (Yun, Donahue, Dudley, & McFarland, 2005), how providing ratings via e-mail versus face-to-face meetings impacts the process (Kurtzberg, Naquin, & Belkin, 2005), and the effectiveness of BARS in measuring team adaptation (Georganta & Brodbeck, 2018). Researchers have argued that no single format is clearly superior to the others across all evaluative dimensions (Landy & Farr, 1980; Murphy & Cleveland, 1995), though a recent paper that reviewed research on BARS makes a case for the superiority of that approach on some dimensions (Debnath, Lee, & Tandon, 2015). I will make no attempt to argue for one approach over another. **Table 5.1** provides a summary of the advantages and disadvantages of the four major types discussed here. You can use this to decide for yourself which format you would use to evaluate your employees; in some cases, your choice would depend on the situation. Note, however, that companies currently often use a graphic rating scale with various behavioral anchors—basically a hybrid of a graphic rating scale and a BARS.

It's also important to note that there has been a good bit in popular press outlets like NPR and the *Huffington Post* about abolishing performance appraisal all together. This suggestion, however, seems to ignore the important uses of performance appraisal in personnel decisions, employee development, and legal documentation, as outlined at the outset of this chapter. Effective procedures in all three of these areas are necessary for successful performance management in organizations. Most organizations value performance appraisal as an important part of the performance management system, but I/O practitioners and researchers alike all agree that there is great room for improvement in appraisal systems and the implementation of those systems.

TABLE 5.1 Summary of Appraisal Formats

	Advantages	Disadvantages
Graphic rating scales	1. Easy to develop 2. Easy to use	1. Lack of precision in dimensions 2. Lack of precision in anchors
BARS	1. Precise and well-defined scales—good for coaching 2. Well received by raters and ratees	1. Time and money intensive 2. No evidence that it is more accurate than other formats
Checklists	1. Easy to develop 2. Easy to use	1. Rater errors such as halo, leniency, and severity are quite frequent
Employee comparison methods	1. Precise rankings are possible 2. Useful for making administrative rewards on a limited basis	1. Time intensive 2. Not well received by raters (paired comparison) or ratees (forced distribution)

Rating Errors

Evaluating another individual's performance accurately and fairly is not an easy thing to do; moreover, errors often result from this process. An understanding of these errors is important to appreciating the complexities of performance appraisal. Research in cognitive psychology has shed considerable light on how the human brain processes information while making decisions. This research has provided I/O psychologists with valuable information that has been applied to the performance appraisal process.

Cognitive Processes In a typical company situation, once or twice a year supervisors have to recall specific performance incidents relating to each employee, somehow integrate those performance incidents into a comprehensible whole, arrive at an overall evaluation, and, finally, endorse a number or category that represents the employee's performance over that period. Furthermore, this has to be done for each of 6 to 12 different employees, and perhaps more! With the trend toward flatter organizations, the number of subordinates for each supervisor is increasing steadily, making the task that much more difficult.

Although more complex *cognitive-processing* models of performance appraisal have been developed (see Hodgkinson & Healey, 2008; Landy & Farr, 1980), all such models are consistent with the scheme depicted in **Figure 5.5.** This figure includes two examples of potential error or bias that may come into play at each of the steps shown (see the left and right columns). The first step in this model is the *observation* of employees' behaviors. In many situations, this is done well; the rater may observe a large portion of the ratee's behavior, as when a grocery store manager observes his subordinates' performance every day. In other situations, however, raters are unable to observe ratees' performance directly. For instance, professors are often evaluated by their department heads, even though the department heads have little opportunity to view the professors engaging in job-related tasks like teaching.

Second, the observed behavior must be *encoded,* which means that the behavior must be cognitively packaged in such a way that the rater is able to store it. If an

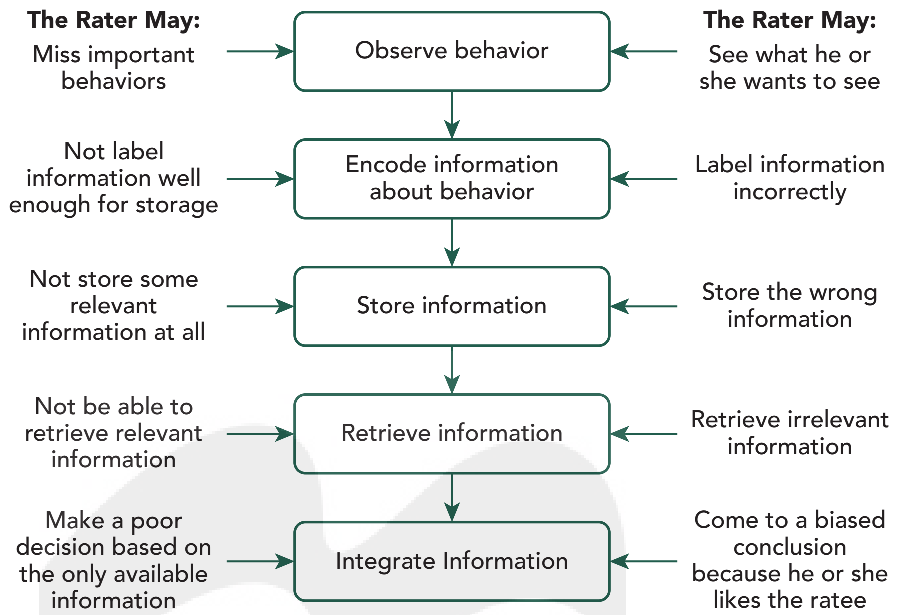
The Rater May:
Miss important behaviors
Not label information well enough for storage
Not store some relevant information at all
Not be able to retrieve relevant information
Make a poor decision based on the only available information
Observe behavior
Encode information about behavior
Store information
Retrieve information
Integrate Information
The Rater May:
See what he or she wants to see
Label information incorrectly
Store the wrong information
Retrieve irrelevant information
Come to a biased conclusion because he or she likes the ratee

FIGURE 5.5 Cognitive-Processing Model of Performance Appraisal

observed behavior is encoded incorrectly (e.g., an adequate behavior is somehow encoded as inadequate), the appraisal rating will be affected at a later time.

Third, after encoding, the behavior must be *stored* in long-term memory. Because it is unreasonable to expect that anyone could perfectly store all the relevant performance incidents, some important incidents may not get stored at all.

Fourth, when the appraisal review is being conducted, the stored information must be *retrieved* from memory. In many situations, the rater cannot retrieve some of the important information, leading to an appraisal rating that is based on an inadequate sample of behavior. Also, since performance reviews are difficult and time-consuming, it's not unusual for raters to retrieve irrelevant information and use it as the basis for the performance rating. Think about doing a review of your subordinate's administrative assistant—someone with whom you have very little interaction. In doing this review, you would certainly get input from your subordinate; but later, when doing the final evaluation, you might recall a memo from your subordinate that explained an expensive mix-up in the office in terms of scheduling conflicts. You may consider this when doing the administrative assistant's evaluation (after all, it makes sense that the assistant would be in charge of scheduling), even though you don't know whether this mix-up was the assistant's fault. For that matter, the assistant may have been the one who caught the problem and saved the company money. If you stored irrelevant information, you may unwittingly use it later in making a performance judgment.

Finally, the rater has to *integrate* all this information and come to a final rating. If the rater has done a good job of observing, encoding, storing, and retrieving relevant performance information, the integration step should be relatively easy. However, sometimes raters let attitudes and feelings cloud their judgments. If your partner, spouse, or even a close acquaintance worked under your direct supervision, could you be objective in arriving at a performance judgment that would be used in a promotion decision? You'd like to think that you could, but many of us probably couldn't. This is one reason that many companies frown on hiring both members of a married couple or, if they do hire them, make sure to limit their workplace interaction. In one case that I know of, the U.S. military accepted a married couple—two attorneys—into the army as captains, guaranteeing that they would be posted to the same geographical region; but the military also guaranteed that the couple would not be assigned to the same base, thereby avoiding potential problems with favoritism and bias.

I/O psychologists have been developing ways to help raters avoid the cognitive errors involved in performance appraisal for the last 35 years. Let's consider some of the most common of these errors.

halo

The rating error that results from either (1) a rater's tendency to use his or her global evaluation of a ratee in making dimension-specific ratings for that ratee or (2) a rater's unwillingness to discriminate between independent dimensions of a ratee's performance.

Halo One error that has received a great deal of attention is called halo. Halo results from either (1) a rater's tendency to use his or her global evaluation of a ratee in making dimension-specific ratings for that ratee or (2) a rater's unwillingness to discriminate between independent dimensions of a ratee's performance (Saal, Downey, & Lahey, 1980). Halo effects can be positive or negative. Personnel managers may give employees who perform well on one dimension higher ratings on another unrelated dimension (Belle, Cantarelli, & Belardinelli, 2017). For example, a retail manager may evaluate a salesclerk as being very good at connecting with potential customers because the manager knows that the clerk is very good at keeping the shelves neat and stocked even though competence on this dimension doesn't really indicate that the clerk is competent in dealing with customers. The manager has generalized the evaluation of this clerk from one dimension to another.

The early research in this area assumed that all halo was error. Traditionally, I/O psychologists noted high correlations across ratings on multiple dimensions—leadership, communication skills, motivating employees, completing paperwork, and so on—and concluded that there was a great deal of halo error. However, some people are competent across *all* dimensions. From a hiring standpoint (see Chapters 6 and 7), organizations tend to target those applicants they believe will be "good at everything." And, indeed, there are employees in all organizations who do seem to be good performers

Sam Diephuis/Blend Images/Getty

A CLOSER LOOK

How can we distinguish between halo and true halo?

on all performance dimensions. The point here is that some halo results from accurate intercorrelations among performance dimensions (see Goffin, Jelley, & Wagner, 2003; Murphy & Jako, 1989; Murphy & Reynolds, 1988)—what we call **true halo**.

true halo
Halo that results from accurate intercorrelations among performance dimensions rather than from rating error.

It is also possible that extremely low intercorrelations among performance dimensions (what some have called negative halo because performance on the dimensions seems to be unrelated) may reflect inaccuracy in ratings just as much as strong intercorrelations (positive halo). In a study that reanalyzed some existing data, a curvilinear relationship between halo and accuracy emerged such that both positive and negative halo were found to reduce accuracy (Thomas, Palmer, & Feldman, 2009). In other words, both very strong and very weak associations between performance dimensions may indicate low accuracy. So, we know that halo exists and we understand it better now than we did 30 years ago, but it is complex. Sometimes halo is reflective of true performance and sometimes either positive or negative halo may reflect inaccuracy in ratings.

TECHNICAL TIP

We discussed the normal distribution in Chapter 2, pointing out that many qualities, such as intelligence and normal personality characteristics, are distributed in a bell-like shape, with most of the data in the middle and much less of the data in the extremes of the distribution.

Performance is one of those qualities that we expect to be distributed "normally." Rating errors such as leniency, central tendency, and severity (discussed next) are categorized as **distributional errors** because they result from a mismatch between actual rating distributions and expected rating distributions. In other words, the grouping of ratings is much farther toward one end of the distribution or much closer to the middle than what we assume the true distribution to be.

distributional errors
Rating errors, such as severity, central tendency, and leniency, that result from a mismatch between actual rating distributions and expected rating distributions.

Leniency Raters commit the error of **leniency** when (1) the mean of their ratings across ratees is higher than the mean of all ratees across all raters or (2) the mean of their ratings is higher than the midpoint of the scale. In other words, if your boss rates employees higher than all the other bosses rate their employees, or if your boss gives ratings with a mean of 4 on a 5-point scale, your boss would be described as a lenient rater. Raters may be lenient because they like their employees or want to be liked. They may think that giving everyone favorable ratings will keep peace in the workplace or that doing so will make them look good as supervisors who have high-performing subordinates. In fact, in a laboratory study examining leniency effects, researchers found that when raters were held accountable to their supervisors (as well as ratees), they were less lenient in their ratings than when they were held accountable only to the ratees (Curtis, Darvey, & Ravden, 2005). This surely has implications for how organizations might want to structure performance appraisal processes.

leniency
The rating error that results when (1) the mean of one's ratings across ratees is higher than the mean of all ratees across all raters or (2) the mean of one's ratings is higher than the midpoint of the scale.

As with halo, we need to be careful in assuming that distributional errors are really errors. Indeed, it is possible, perhaps likely, that some supervisors really do have better employees or work groups than others, resulting in more favorable ratings that are accurate rather than lenient. For example, the airline chosen as the best in the nation with respect to service should show higher performance ratings for its service employees than other airlines. So what might appear to be the result of leniency among the evaluators of this airline would in fact be the result of accurate evaluation. Research also suggests that personality can have an impact on one's tendency to be lenient. Specifically, individuals categorized as "agreeable" have been shown to be more lenient than those categorized as "conscientious" (Bernardin, Tyler, & Villanova, 2009). In a more recent study, raters who were high on agreeableness and extraversion tended to rate others most leniently (Cheng, Hui, & Cascio, 2017).

Central Tendency Raters who use only the midpoint of the scale in rating their employees commit the error of **central tendency**. An example is the professor who gives almost everyone a C for the course. In some cases, the raters are lazy and find it easier to give everyone an average rating than spending the extra time necessary to review employees' performance so as to differentiate between good and poor workers. In other cases, the raters don't know how well each of their subordinates has performed (perhaps because they are new to the work group or just don't see their employees' on-the-job behavior very often) and take the easy way out by opting for average ratings for everyone. Some research suggests that central tendency error is sometimes a result of the rating scale itself; relatively simple semantic differential scales (e.g., ranging from "effective employee" to "ineffective employee") result in a considerable amount of this bias (Yu, Albaum, & Swenson, 2003).

central tendency The tendency to use only the midpoint of the scale in rating one's employees.

Of course, a given work group or department may be populated largely by average employees. In fact, the normal distribution suggests that most employees really are average, so it is reasonable to have a large percentage of employees rated as such. At times, then, central tendency is a rating error; at other times, though, it simply reflects the actual distribution of performance, which is largely centered around "average."

Severity Less frequent than leniency and central tendency is the rating error of **severity**, which is committed by raters who tend to use only the low end of the scale or to give consistently lower ratings to their employees than other raters do. Some supervisors intentionally give low ratings to employees because they believe that doing so motivates them (you will see when we get to Chapter 9, on motivation, that this strategy is not likely to work), or keeps them from getting too cocky, or provides a baseline from which new employees can improve. For some, severity represents an attempt to maintain the impression of being tough and in charge—but what tends to happen is that such raters lose, rather than gain, the respect of their subordinates.

severity The tendency to use only the low end of the scale or to give consistently lower ratings to one's employees than other raters do.

Some work groups include a larger number of low performers than other work groups, so low ratings from the supervisor of such work groups may be accurate rather than "severe." Thus, although we don't see these terms in the literature, we could speak of true leniency, true central tendency, and true severity in much the same way as we speak of true halo. For any given situation, though, it is difficult to determine whether the ratings are affected by rating errors or are an accurate reflection of performance.

The chief problem stemming from distributional errors is that the ratings do not adequately *discriminate* between effective and ineffective performers. In such cases, the majority of ratees are lumped together in the bottom, middle, or top of the distribution. This general problem is often referred to as *range restriction* because only a small part of the scale range is used in the ratings. The difficulty for the organization is that it intends to use performance rating information for personnel decisions such as promotions, raises, transfers, layoffs, and other terminations; but if all the employees are rated similarly (whether as a result of central tendency, leniency, or severity), the ratings do not help in making these personnel decisions. If everyone is rated in the middle of the scale, who gets promoted? If everyone is rated very favorably, who gets the big raise? If everyone is rated as ineffective, who gets fired?

Employee morale is also affected by ratings that don't discriminate on the basis of performance, in that employees who believe they are good employees will feel slighted because their reviews are no better than those of employees whom they view as much less effective. Think about this: Have you ever been passed over for a promotion, only to discover that the person who was promoted, though less deserving than you, was rated similarly to you? You likely experienced feelings of *injustice,* which may have affected not only your subsequent on-the-job performance but also your attitude. These additional implications of nondiscriminating ratings will be discussed in more detail in Chapters 10 and 11.

Other Errors Many other rating errors are discussed in the literature, but I will touch on just three more. One is *recency error,* whereby raters heavily weight their most recent interactions with or observations of the ratee. A supervisor who, in rating his subordinate, largely ignores nine months of superior performance and bases his evaluation on only the past three months of less-than-adequate performance is making a recency error. This is similar to the somewhat misguided belief in organizations that all that matters is the question "What have you done for me lately?"

Another error of note, called *first impression error* or *primacy effect,* is the opposite of recency error. Here, raters pay an inordinate amount of attention to their initial experiences with the ratee. A construction foreman may think back to that first day when the new electrician helped out on the site at a crucial time and use this as the basis for his evaluation of the electrician while largely ignoring some major performance problems over the past few months. First impressions tend to be heavily weighted in our everyday lives, as when we form friendships with people with whom we just seem to "hit it off" from the very beginning; they are also used, sometimes ineffectively, in performance appraisal.

Finally, there is the *similar-to-me* error, which occurs when raters tend to give more favorable ratings to ratees who are very much like themselves. We know from social psychology that people tend to make friends with and like being around people who are much like themselves. An old English proverb states: "Birds of a feather flock together." A similar effect occurs in performance appraisal situations, resulting in more favorable ratings of employees similar to the rater than of those dissimilar.

I/O TODAY

Should We Eliminate Performance Ratings?

At a SIOP conference in 2015, a "standing-room-only" debate occurred between researchers and academics over whether getting rid of performance ratings completely was a good idea, or a recipe for disaster (Adler et al., 2016). Indeed, almost nobody enjoys the performance appraisal process, and many business publications have suggested eliminating it altogether (e.g., Resker, 2017; Ryan, 2018). What are the arguments for eliminating performance ratings, and do they have any merit? First, evidence shows that employees do not like receiving feedback, and often are so demotivated after receiving feedback that their performance actually decreases (DeNisi & Smith, 2014). This typically occurs because employees focus on the ratings rather than strategizing on how to change their behaviors to improve their performance.

As you read through this chapter, you will see that there are many problems with rating scales, such as leniency and halo, that prevent organizations and employees from trusting the information provided by performance ratings. There are a number of contextual factors that might affect how an employee is rated, including the labor market, team dynamics, leadership, and organizational culture. There are also political concerns around appraisal ratings—for example, if a manager often gives negative ratings to her team members, she may be seen as difficult or as an ineffective mentor to her team, which might result in her getting fired or receiving fewer resources from the organization.

Based on these arguments, eliminating performance *ratings* might appear to be a reasonable strategy, but in actuality eliminating performance evaluations is probably not a great idea. Organizations must have some way of evaluating and quantifying employee performance in order to defend personnel decisions. Research also suggests that strong performers are more likely to join and remain in organizations where their hard work is recognized (Allen & Griffeth, 2001; Menefee & Murphy, 2004). There is no clear single solution here, but scholars and practitioners have provided some guidelines for finding better ways to evaluate employees. Waters, Braughman, and Dorsey (2016) emphasized the value of more frequent performance discussions (as opposed to annual reviews) that focus on helpful, tangible outcomes. Scholars have also emphasized that part of the problem with performance appraisals is that managers often view them as the culmination of the feedback process rather than the beginning—the feedback is useless unless employees receive coaching, mentoring, and training to help them improve (Rotolo et al., 2018). Companies including Adobe, Cargill, and Gap have eliminated annual performance reviews and replaced them with check-ins, during which managers can correct problematic behaviors quickly and reward hard workers with higher salaries (Rotolo et al., 2018). Performance evaluation will continue to evolve over the next decade, and I/O psychologists will need to be agile in order to respond to innovations and find new and better ways to assess employee performance.

Discussion Questions

1. Consider a time when you received negative feedback about your performance on a task (on a work task or a nonwork task, such as playing an instrument). Were you able to successfully improve your performance? What type of feedback did you find helpful or do you wish you would have received?
2. Imagine you receive feedback very infrequently (perhaps once per year, or once per semester). In what ways would that make your job (or studies) more difficult? On the other hand, do you think it's possible to receive feedback too frequently?
3. Have you experienced or heard about any unique approaches to evaluating and/or rewarding performance? What might be some advantages and disadvantages of that approach?

Rater Considerations

We've talked at length about characteristics of the rating—that is, formats and errors. In this section, we consider important performance appraisal elements that revolve around the rater—namely, rater training, rater goals, and rater accountability.

Rater Error Training (RET)
A type of training originally developed to reduce rater errors by focusing on describing errors like halo to raters and showing raters how to avoid making such errors.

Rater Training We have just discussed some of the common errors that raters make in evaluating the performance of others. An important question asked by I/O researchers and practitioners alike is whether rater training can reduce such errors and improve the rating process (Hauenstein, 1998; Schleicher, Day, Mayes, & Riggio, 2002). There are two main types of rater training in the performance appraisal area. One, known as **Rater Error Training (RET)**, was originally developed to reduce the incidence of rater errors (Spool, 1978). The focus was on describing errors like halo to raters and showing them how to avoid making such errors. The assumption was that by reducing the errors, RET could increase accuracy, the degree to which performance ratings match one's true performance level.

As suggested early on in the development of this approach, however, RET can indeed reduce errors, but accuracy is not necessarily improved. In fact, studies have shown that accuracy sometimes *decreases* as a function of reducing error (e.g., Bernardin & Pence, 1980). How can this be? Well, recall our discussion of halo. When raters are instructed not to allow for halo, they are in effect being taught that there is no relationship among performance dimensions and that their ratings across dimensions should not be correlated. But, in many cases, there is true halo, and the ratings *should* be correlated. Thus, rater training may have reduced the correlation across dimensions that we used to assume was error, resulting in artificially uncorrelated ratings that are now inaccurate.

Frame-of-Reference (FOR) training

A type of training designed to enhance raters' observational and categorization skills so that all raters share a common view and understanding of performance levels to improve rater accuracy.

A second type of rater training, called **Frame-of-Reference (FOR) training**, was designed by John Bernardin and his colleagues to enhance raters' observational and categorization skills (Bernardin & Beatty, 1984; Bernardin & Buckley, 1981). Bernardin's belief was that to improve the accuracy of performance ratings, raters have to be provided with a frame of reference for defining performance levels that is consistent across raters and consistent with how the organization defines the various levels of performance. In other words, for a fast-food employee, on the dimension of cleanliness, all raters must know that on a 5-point scale a 5 would be indicated by the following behaviors:

> Tabletop is wiped down in between runs of burgers. Ketchup and mustard guns are always placed back into their cylinders and never left sitting on the table. Buns are kept wrapped in plastic in between runs of burgers. Floor around production area is swept at least once per hour and mopped once in the morning and afternoon. All items in the walk-in refrigerators are placed on the appropriate shelves and the walk-ins are organized, swept, and mopped periodically.

FOR training attempts to make that description part of all raters' performance schema for the level of "5/exceptional" performance. The hope is that by etching this performance exemplar in the raters' minds, the training will render each rater better able to use it consistently when observing, encoding, storing, retrieving, and integrating behaviors in arriving at a final rating. The goal is to "calibrate" raters so that a score of 5 from one rater means the same as a score of 5 from any other rater. Popular procedures in FOR training have since been developed (Pulakos, 1984, 1986). In these, raters are provided with descriptions of the dimensions and rating scales while also having them read aloud by the trainer. The trainer then describes ratee behaviors that are representative of different performance levels on each scale. Raters are typically shown a series of videotaped practice vignettes in which individuals (stimulus persons, or ratees) are performing job tasks. Raters evaluate the stimulus persons on the scales; then the trainer discusses the raters' ratings and provides feedback about what ratings should have been made for each stimulus person. A detailed discussion ensues about the reasons for the various ratings.

Antonio_Diaz/Getty Images

A CLOSER LOOK

How can FOR training eliminate some of the rater errors, like primacy and recency error, discussed earlier?

Research has consistently shown that FOR training not only improves appraisal accuracy (Sulsky & Day, 1994) but also is generally recognized as the most effective approach for improving rater accuracy (Meriac, Gorman, & Macan, 2015). A couple of studies have suggested that combining FOR training with *behavioral observation training (BOT),* which focuses on teaching raters how to watch for certain behaviors and avoid behavioral observation errors, may improve the recognition or recall of performance behaviors (Noonan & Sulsky, 2001; Roch & O'Sullivan, 2003). One current concern, however, is that very little field data are available on the FOR technique. Most of the research showing that FOR training improves accuracy has been conducted with students in laboratory settings (see Uggerslev & Sulsky, 2008). For instance, in a recent meta-analysis that found strong moderate effects of FOR training on various accuracy measures, only 2 of the 39 manuscripts identified since 1994 described research studies that did not use students (Roch, Woehr, Mishra, & Kieszczynska, 2012). There were more studies in this literature review that did use FOR training in field settings, but these studies were not direct tests of the effectiveness of FOR training and could not be included in the meta-analysis. The next step is to find a way to bring FOR training research into more organizational settings, but doing so requires some modification to the rather expensive and time-consuming process. Some recent attempts to bridge this gap have been successful, such as one study that used diverse experimental designs in the lab and field and showed the positive impact of FOR training for both students and workers (Moser, Kemter, Wachsmann, Köver, & Soucek, 2018).

Rater Goals and Accountability The effect of rater goals/motivation, along with accountability related to the appraisal process, is also quite important and has been the focus of considerable research. Raters may have different goals for various reasons. For instance, a particular rater may want all of his or her subordinates to be rated above a certain level or to have developed in particular areas. These kinds of goals or objectives may affect the ratings provided or the rating process in general. Sometimes this is reflected in raters "bending the rules" or making performance-based judgment calls to enhance individual or organizational performance (Veiga, Golden, & Dechant, 2004); at other times, these goals are reflected in how the rating process is implemented.

Mero and his colleagues have demonstrated that raters who were held accountable to various goals or objectives, such as rating accurately or rating leniently, actually provided ratings consistent with those goals (Mero & Motowidlo, 1995). One step further, it was shown that certain types of raters are more affected by accountability pressures than are others (Mero, Guidice, & Anna, 2006). For instance, raters who are high on conscientiousness tend to be strongly affected by accountability pressures, which is reflected in a felt need to justify one's ratings, some anxiety about the task, and higher-quality performance ratings.

These researchers have also found that the accountability effect was altered as a function of the status of the audience and the manner of being held accountable (Mero, Guidice, & Brownlee, 2007). For example, raters who expected to be held accountable for their ratings in a face-to-face meeting with a high-status session administrator provided more accurate ratings than those who anticipated justifying ratings only in writing. Finally, Wong and Kwong (2007) demonstrated a clear relationship between rater goals and rating patterns. When raters were asked to maintain harmony within the work group, they increased their mean ratings and tended not to discriminate between rates.

CONTEMPORARY PERFORMANCE APPRAISAL RESEARCH

For many years, practitioners and researchers believed that rating-scale formats were integral to the success of performance appraisal systems and that *accuracy* should be the chief goal of any performance appraisal system. Because of these two common beliefs, much of the early research in this area focused on the context of performance appraisal as it related to rating formats and rater errors. Format and error research has taught us a great deal about the processes and errors involved in arriving at a performance judgment. However, performance appraisal is not a stand-alone process that can be examined in isolation but, rather, is a complex process that takes place in a very rich and sophisticated social-psychological climate, or **context**. Hence, the remaining sections of this chapter will focus on contemporary issues that better emphasize that context (see Levy, Cavanaugh, Frantz, Borden, & Roberts, 2018; Levy & Williams, 2004).

context
The social-psychological climate in which performance appraisal takes place.

The Importance of the Social-Psychological Context

Many experts have suggested a more direct research focus on the context in which the appraisal takes place, arguing that this context colors the entire appraisal process. The context in which performance appraisal takes place includes not only the social and legal climates in which the organization exists but also the political and emotional climates and cultures within the organization itself (Ferris, Munyon, Basik, & Buckley, 2008).

These context-related topics include an examination of (1) the use of employees' reactions to the appraisal rather than *accuracy* as a criterion for evaluating performance appraisal systems, (2) how the relationship between the supervisor and subordinate affects performance appraisal, (3) the role of organizational politics in the appraisal process, (4) the importance of trust in the appraisal process, (5) the use of multiple feedback sources rather than just the supervisor's feedback in the appraisal process, and (6) the value of providing employees with knowledge about the appraisal system and the opportunity to participate in the appraisal process. This list represents just a small sampling of the context issues at the core of current performance appraisal research (see Levy et al., 2018, for a complete review). With our movement away from an emphasis on formats and errors, we are beginning to increase our understanding of the process itself and to apply that understanding to the use of appraisals in organizations. The next few sections will consider some of this new context-based research.

Reaction Criteria Traditionally, performance appraisals were evaluated with respect to how "accurate" they were, in the sense of being free from errors. But as you now know, there are problems with accuracy and error measures. As early as 1984, researchers suggested that future appraisal research should begin to move beyond psychometric criteria due to the measurement problems inherent in that approach and instead focus on *qualitative criteria* (Bernardin & Beatty, 1984) and *reaction criteria* (Cawley, Keeping, & Levy, 1998). Both phrases refer to the role played by raters' and ratees' reactions in the appraisal process (Hedge & Teachout, 2000; Keeping & Levy, 2000).

One research focus has been the potential importance of raters' and ratees' favorable responses to the appraisal system or process (e.g., Findley, Giles, & Mossholder, 2000). In fact, until the last 15 years, the relative lack of research attention directed toward reaction criteria instead of psychometric and accuracy criteria led researchers to refer to reaction criteria as one class of "neglected criteria" that might be critical in evaluating the success of an appraisal system (Murphy & Cleveland, 1995). It seems reasonable to expect that subordinates' reactions to appraisal systems would have just as much impact on the success and effectiveness of an appraisal system as the more technical aspects of the system. Indeed, researchers have suggested that reactions are usually better indicators of the overall viability of an appraisal system than are narrower psychometric indices such as leniency or halo (Bernardin & Beatty, 1984). After all, one may develop the most technically sophisticated appraisal system, but if that system is not accepted and supported by employees, its effectiveness will ultimately be limited. Indeed, due to the changing nature of performance appraisals and organizations, worker reactions toward performance appraisal may play an increasingly important role in appraisal processes as the procedures and systems continue to develop (Hedge & Borman, 1995).

Despite the relative neglect of reaction criteria, several studies have attempted to investigate various appraisal characteristics that elicit or at least contribute to positive employee reactions. For instance, it has been demonstrated that for a sample of bank tellers, appraisal satisfaction was affected by both appraisal characteristics (e.g., frequency of appraisals) and organizational variables (e.g., employees' understanding of their jobs). Furthermore, bank tellers who reported a positive relationship with their supervisors and a low amount of ambiguity about their roles in the organization were more satisfied with the appraisal system (Dobbins, Cardy, & Platz-Vieno, 1990).

One study demonstrated that subordinates reported less anger and higher perceptions of justice when supervisors provided justification for their ratings (Tata, 2002). Furthermore, we have seen that performance appraisals viewed as unfair by subordinates have tended to be related to ratees' emotional exhaustion (Brown & Benson, 2003). (We will talk more about organizational justice in a later section of this chapter and in Chapters 9 and 10.) This research has great implications for appraisal in organizations because it suggests that organizations and supervisors can do certain things to increase the likelihood that employees will respond favorably to the performance appraisal process—namely, provide frequent performance feedback and relevant job-related information to employees. In a recent conceptual paper, affect has been integrated into performance management as something that is experienced after one receives his or her rating and also something that impacts the way raters rate—this conceptualization broadens the role of affect in performance appraisals (Roberts, Levy, Flores, & Thoebes, in press).

A recent study modeled the goals-feedback-reactions-subsequent-goals path and determined that emotional reactions to feedback play a significant role in determining future goals (Ilies, Judge, & Wagner, 2010). The authors argue that emotion and cognition are instrumental in determining ensuing goals and that this process has important organizational implications for performance as evidenced through the established literature on goal setting (see Chapter 9 for further discussion on goal-setting theory).

Justin Kruger and David Dunning (1999) have done some very interesting research arguing that there are people who are so unskilled that they aren't skilled enough to know how unskilled they are. The Dunning-Kruger effect is a cognitive bias that basically suggests that unskilled individuals overestimate their abilities and do not have the meta-cognitive skills to even realize that they are doing this. As you can imagine, this lack of judgment can cause a variety of problems for individuals and organizations when it comes to properly setting and evaluating goals. Some of Dunning and colleague's more recent work has extended this to the feedback arena, where they show that individuals who are unskilled and unaware are also uninterested in learning and getting better (Sheldon, Dunning, & Ames, 2014). These researchers found that those individuals who were low on *emotional intelligence (EI)*—an individual's ability to deal effectively with his or her emotions and the emotions of others (see Chapter 6)—also tended to overestimate their EI (Dunning-Kruger effect). Interestingly, when given feedback and the opportunity to use that feedback to improve, these individuals were unlikely to do so. More specifically, they tended not to drop their self-ratings of EI much even after feedback; they tended to evaluate the test as poor and inaccurate; and they showed less interest in self-improvement, which was measured by, among other things, their interest in purchasing a book called *The Emotionally Intelligent Manager.* Those low on EI were much less interested in purchasing the book and when they did express an interest in purchasing the book, they were willing to spend far less than those who were high on EI. They were, as Sheldon and colleagues (2014) noted, unskilled, unaware, and uninterested in learning! The results from this line of research are consistent with those that practitioners often tend to see in organizations: The people who need feedback the most either don't get it or are not receptive to it, but rather react to it in a negative, disparaging way. This tendency complicates the feedback process, which relies on managers who will work hard to create an atmosphere where employees hear, receive, accept, and use feedback to improve. A recent paper makes the following recommendations to improve reactions to feedback: Provide more regular feedback sessions, invest in positive manager-subordinate relationships, rethink how ratings are done, and promote fairness/organizational justice (Levy et al., 2018).

Digital Vision./Getty Images

A CLOSER LOOK

The relationship between supervisor and subordinates, sometimes called LMX, is very important for organizational functioning. How do theory and research related to LMX help us understand the performance appraisal process?

The Supervisor–Subordinate Relationship

Researchers and practitioners have discussed the relationship between supervisors and subordinates in terms of *leader–member exchange (LMX) theory,* which emphasizes the idea that supervisors have different types of relationships with different subordinates (Engle & Lord, 1997; see Chapter 13 for a more detailed discussion). In short, the relationship that an employee has with his or her supervisor plays a role in the performance appraisal process. A recent meta-analysis supports this conclusion, uncovering a .34 correlation between LMX and performance (Martin, Guillaume, Thomas, Lee, & Epitropaki, 2016). Further, the meta-analysis also found that this relationship was mediated by other factors, with trust and empowerment being the strongest. That is, individuals who reported that they had a positive relationship with their bosses also reported high trust and empowerment. These same individuals were also rated favorably by their supervisors with respect to performance.

An interesting study demonstrated that the frequency of communication between a subordinate and supervisor interacts with LMX to affect performance ratings. A positive LMX relationship led to favorable ratings when there was also frequent communication, but ratings were less favorable even with a positive LMX when there was little communication (Kacmar, Witt, Zivnuska, & Gully, 2003). This is an excellent example of the connection between the context in which the appraisal takes place and the outcome of the appraisal process. The implication in this case is that understanding the appraisal process within a work group requires an understanding of the relationships among those involved because these relationships affect the appraisal process.

In a recent study of about 220 supervisor–subordinate dyads in Dubai, the authors looked at three of the constructs we have been talking about in this section: LMX, justice, and performance reactions (Pichler, Varma, Michel, & Levy, 2015). They found that the quality of the LMX relationship affected performance appraisal satisfaction through its effect on procedural justice. In other words, if subordinates had a positive relationship with their supervisor, they tended to perceive that they were being treated more fairly and this was reflected in a more favorable evaluation of the appraisal process.

Park (2017) presents a mode in which LMX quality is proposed to result in more rater accountability due to high levels of interaction, feedback discussions, and information sharing. Another recent conceptual paper (Tseng & Levy, 2018) recommends a focus on manager–employee interactions as a way to improve the effectiveness of performance management. This innovative framework highlights the way in which managers impact this process through their interactions with employees, teams, and the organization. These new conceptual approaches promise to drive future research in this area.

Organizational Politics Traditionally, organizational politics has been defined as "deliberate attempts by individuals to enhance or protect their self-interests when conflicting courses of action are possible" (Longenecker, Sims, & Gioia, 1987, p. 184). This description can help delineate the context in which performance appraisal takes place. Because organizations are political entities, politics plays a role in all important organizational decisions and processes, including performance appraisal (Longenecker et al., 1987). It's important to note that both supervisors and subordinates are affected by the politics of performance appraisal. In particular, Clint Longenecker and his colleagues (1987) have demonstrated through interviews that executives intentionally manipulate appraisals for political reasons (e.g., to send a message to subordinates or to make their own department look good). Research has also indicated that subordinates use *impression management* strategies in an effort to bring about favorable appraisals; in other words, they control their behavior to make a good impression on their superiors (Wayne & Liden, 1995).

Recent research has begun to examine how perceptions of politics relate to constructs like stress, anxiety, and morale within a performance context. For instance, one study uncovered relationships among perceptions of politics, attitudes, stress, and performance (Rosen & Levy, 2013). In particular, the researchers found that when employees perceive politics at work they also tend to experience stress and anxiety, which impact their work attitudes and result in performance decrement.

In another study, researchers looked at the effect of politics perceptions and rumination on performance (Rosen & Hochwarter, 2014). *Ruminators* are individuals who persist in negative thinking and feeling related to stressors that they experience. In studies of architects, clerical employees, and accountants, researchers found that perceptions of politics operated differently depending on an individual's tendency to ruminate. In particular, they found that politics had a strong negative effect on performance (as well as job satisfaction, tension, and mood) for ruminators, but a much weaker effect for those who tended not to dwell on stress-related forces. This finding suggests that politics may not be so detrimental to employees who manage not to ruminate on the situation.

A recent review and look to the future presents organizational politics as a neutral construct in that there are both positive and negative elements of it (Ferris, Ellen, McAllister, & Maher, 2019). The researchers define politics as the management of shared meaning, expanding the traditional conceptualizations to include political characteristics, political actions, and political outcomes. Each of these aspects of politics has various dimensions, but chief among them are political skill, political behavior, and perceptions of politics, respectively. Further, their review of the literature suggests potential benefits, such as higher ratings for those individuals who are politically skilled.

Finally, in a simple but interesting study, researchers observed college interns and found a revealing pattern of results (Liu, Ferris, Xu, Weitz, & Perrewe, 2014). They focused on ingratiation, which is an influence tactic that uses behaviors to curry favor or make oneself look good to others—some of these behaviors are very similar to "kissing up." Liu and colleagues (2014) found that interns who use ingratiation tactics are rated as better performers by their supervisors if they are also politically skilled. In other words, if an individual is skilled at ingratiation, it works—his or her supervisor will like the intern and this will be reflected in positive ratings. However, for those interns who used ingratiation tactics and were not good at it, this tactic resulted in low performance ratings. So, the world of politics, ingratiation, and performance is complicated, but I'm guessing you already knew that. Now you have science to support it!

Trust and Justice *Trust* in the appraisal process is the extent to which raters believe that fair and accurate appraisals have been or will be made in their organization. Alternatively, if a rater feels that other raters in the organization are inflating their ratings, that rater may do the same. Bernardin and his colleagues developed a scale called Trust in the Appraisal Process Survey, or TAPS, to measure raters' perceptions of the rating behavior of the other raters in their department (Bernardin, 1978; Bernardin, Orban, & Carlyle, 1981). They found that raters who scored low on the TAPS were more lenient raters than those who scored high. This finding indicates the importance of trust on appraisal ratings. In addition, performance appraisal systems that are well received by employees appear to affect employees' trust of top management, suggesting that the performance management system may be an effective tool for enhancing organizational trust (Mayer & Davis, 1999).

In another line of work related to trust issues, researchers have looked at how comfortable raters are with the performance appraisal process (Villanova, Bernardin, Dahmus, & Sims, 1993). In particular, Peter Villanova and his colleagues have shown that scores on a scale they call the Performance Appraisal Discomfort Scale (PADS) are related to leniency in such a way that those raters who express great discomfort in evaluating others and providing them with feedback also tend to be among the most lenient raters. In other words, not all raters are equally comfortable doing performance appraisals, and the extent to which raters experience discomfort in this setting is likely to affect the quality of the ratings and other elements of the appraisal process (e.g., interpersonal interactions). In their more recent work, Bernardin and Villanova (2005) have taken an important additional step by showing that raters can be trained to be more comfortable doing performance appraisal, resulting in less leniency. Called self-efficacy training, this approach has begun to influence rater training, though more research is needed to evaluate its effectiveness.

Jetta Productions Inc/Getty Images

A CLOSER LOOK

How do rater comfort and ratee trust impact the appraisal system?

We will talk about organizational justice in Chapter 9, but the notion of the fairness or justice of a performance appraisal system is an important one to mention here. A carefully designed, psychometrically sound performance appraisal system has the potential to be effective, but that potential can be limited by the justice perceptions of those organizational members who are involved in the appraisal process. If raters or ratees see the appraisal system as unfair or biased, its psychometric quality will be irrelevant—it won't matter how "good" the system is if people perceive it to be unfair.

Recent conceptual work argues that performance appraisal research must be broad in its focus and include diverse aspects of the appraisal system, such as justice-related constructs (Thurston & McNall, 2010). An interesting recent line of work has looked at how a supervisor's *implicit person theory (IPT)*—the extent to which an individual believes that people can change—impacts performance appraisal. One such research investigation (Heslin & VandeWalle, 2011) demonstrated that supervisors who tend to believe that people can change and develop are perceived by their subordinates as more just in their performance appraisals than are those supervisors who do not believe that people can change. The justice perceptions led to more frequent organizational citizenship behaviors (OCBs) and greater commitment to the organization (see Chapter 10 for more on OCBs and organizational commitment). Another interesting application of justice to performance appraisal is found in a study linking negative feedback (feedback indicating subpar performance) to organizational justice (Chory & Westerman, 2009). In their study of 250 working adults, the authors found that the quality of the negative feedback—for instance, the extent to which it was destructive versus constructive or inconsistent versus reliant on clear standards—was strongly related to perceptions of justice. They concluded that managers should focus on providing consistent and constructive feedback to elicit perceptions of justice from employees, hopefully resulting in the feedback being deemed as useful and therefore more likely to be put into practice to improve employees' performance. There is a sizable amount of literature on this (for a more thorough discussion, see Elicker, Levy, & Hall, 2006; Thurston & McNall, 2010), but I want to emphasize the important role played by organizational members' perceptions of performance appraisal fairness.

Participation Employee participation is a contextual process variable that has received a great deal of attention in the performance appraisal literature. Overall, research suggests an association between allowing employees to participate in the appraisal process and positive employee reactions toward the appraisal system (Cawley et al., 1998). Specifically, as demonstrated by a meta-analysis, there is a strong positive relationship between participation and a host of reactions to performance appraisal, including (1) satisfaction with the appraisal system and session, (2) motivation to improve as a result of the appraisal, (3) belief in the fairness of the appraisal, and (4) belief in the usefulness of the appraisal (Cawley et al., 1998). In short, when employees are allowed to participate in the appraisal process—by completing a self-assessment or expressing their ideas during the appraisal session or interview—they react more positively to the appraisal than when they are not given the opportunity to participate. Here, too, the important

TECHNICAL TIP

Remember from Chapter 2 that meta-analyses are quantitative reviews of research findings. We use them to draw useful summaries of relationships among variables from many published and unpublished studies.

implication is that organizations and supervisors have interventions available to them that can improve the quality of the appraisal process—in this case, by giving employees a voice in the process.

Table 5.2 presents the steps to be followed during a performance appraisal interview. Compiled from various sources (primarily Silverman, 1991), these steps make it very clear that the performance appraisal interview should be a participative process

TABLE 5.2 How to Conduct a Performance Appraisal Interview

Prior to the interview	During the interview	After the interview
1. Supervisor should give employee adequate notice about meeting date.	1. Supervisor should explain the purpose of interview.	1. Supervisor and employee should independently review major responsibilities and objectives.
2. Employee should do a self-appraisal.	2. Employee should summarize accomplishments and needs with respect to major responsibilities.	2. Supervisor and employee should schedule additional meetings to set objectives for the next review period.
3. Supervisor should receive copy of self-appraisal.	3. Supervisor should do same summary as above, but from his or her viewpoint.	3. Supervisor and employee should agree on the organizational link between achieving objectives and receiving organizational rewards (compensation, promotions, etc.).
4. Supervisor should review documentation of performance.	4. Supervisor and employee should discuss whether a developmental need exists with respect to each major responsibility.	4. Communication channels between supervisor and employee should remain open throughout the review period.
	5. Supervisor and employee should diagnose the causes of any discrepancy, or "gap," between objectives and actual performance.	
	6. Supervisor and employee should develop action plans to enhance performance on each major responsibility.	
	7. Supervisor and employee should summarize performance on each major responsibility and review agreed-upon action plans.	
	8. Supervisor should compliment employee on accomplishments.	
	9. Supervisor should set time and date for future meetings to discuss responsibilities and performance.	

that includes substantial input from the employee. This observation is consistent with findings regarding the importance of participation in the performance appraisal process (Cawley et al., 1998), as well as with the change in management style and structure from autocratic to humanistic that has become pervasive in organizations (Meyer, 1991).

Providing Performance Feedback

The performance management cycle is a multistage, longitudinal process that is central to the development of employees and the success of organizations (London & Smither, 2002). The key element to this cycle is the extent to which employees receive and use feedback. This focus on employee development has received a great deal of attention. For instance, London and Smither (2002) have presented a comprehensive model of the performance management process in which the chief outcomes are behavior changes, better performance, increased self-awareness, and increased self-confidence. They propose that these outcomes are affected by constructs such as the employee's own feedback orientation (see Linderbaum & Levy, 2007) or receptivity to feedback, as well as the feedback environment as defined by the organization or work group (see Whitaker, Dahling, & Levy, 2007). In other words, employees' development is largely a function of their receptivity to feedback and the organization's approach to or emphasis on feedback (Gabriel, Frantz, Levy, & Hilliard, 2014).

Continuous employee development has been defined as a "cyclical process in which employees are motivated to plan for and engage in actions or behaviors that benefit their future employability on a repetitive or ongoing basis" (Garofano & Salas, 2005, p. 282). One study demonstrated that how participants responded to employee development goals was affected by their personalities (core self-evaluations, which are related to self-efficacy and self-esteem as well as other traits) and other situational and interpersonal variables (Bono & Colbert, 2005).

As noted earlier, feedback accountability plays an important role in both performance management and, in particular, employee development. Research has found that social support is a key determinant of attitudes toward employee development and that organizational social support for development is likely to lead to a feeling of accountability on the part of the employees (Maurer, Mitchell, & Barbeite, 2002). Supervisors who have experienced the support of their subordinates have reported more feedback accountability and greater self-development initiative, suggesting that the social context is an important element of employee self-development (Rutkowski & Steelman, 2005). Let's talk some more about the feedback process.

The Feedback Process Feedback serves an important role in organizations, meeting the needs of employees and employers. London and Smither (2002) highlight the key steps of the performance management cycle. First, employees *attend to the feedback,* which includes anticipating, receiving, and reacting to the feedback. After this initial interaction with the feedback comes the second stage, *processing the feedback,* which includes interpreting, understanding, dealing with, and believing or discounting the feedback. Finally, the last step is *using the feedback,* which involves using the feedback to set goals and to track progress. These three steps are integral to making effective use of feedback; of course, if any of the three steps are ignored or handled ineffectively, this will likely result in feedback-related problems such as misperceptions, anxiety, a lack of understanding, and unwise behavioral choices.

In addition to the important steps of the performance management cycle, we also need to consider qualities or characteristics of the feedback itself if we are to truly understand the process. For instance, what makes for good feedback? What kind of feedback is most effective? A very important study by Kluger and DeNisi (1996) created a stir by arguing that feedback is not as effective and beneficial as we had long thought. They did a meta-analysis of research on feedback interventions and, while showing a positive moderate effect of feedback on performance ($r = .41$), the analysis also revealed that over one-third of the studies showed negative effects and many found no effect at all. The authors fit these results into the historical literature on feedback and argued that we had been misled by this historical literature. They consequently proposed the *feedback intervention theory (FIT),* which states that feedback is most effective when it is targeted at the task rather than at the self. They have argued that feedback directed anywhere other than to the task (e.g., "You are such a conscientious employee") distracts the individual and pulls resources and attention from the task at hand, which leads to no change in performance or a decrease in performance. This is a popular theory and Kluger and DeNisi (1996) have provided some empirical support.

Feedback Environment and Feedback Orientation The culture of an organization serves to constrain or enhance the effectiveness of feedback. The term **feedback environment (FE)** has been proposed to embody that culture while also specifically representing an organization's climate and attitude toward feedback. Research has shown that FE includes such dimensions as source credibility, feedback quality, and feedback delivery, among others. A favorable FE has been linked to satisfaction with and motivation to use the feedback (Steelman, Levy, & Snell, 2004) as well as an interest in seeking feedback. In addition, a positive FE has been tied to higher levels of perceived fit, commitment, and engagement in OCBs, in addition to less burnout (Peng & Chui, 2010). Another study found that supervisor FE was linked to LMX, which positively impacted engagement in OCBs and negatively impacted deviance behaviors (Peng & Lin, 2016).

feedback environment (FE)
The contextual aspects of the day-to-day supervisor–subordinate and coworker–coworker feedback process.

Whereas FE is the situational side of the feedback processes, the **feedback orientation (FO)** is the individual-difference side. FO is an individual's overall attitude toward feedback or receptivity to feedback. Its dimensions include perceptions of the utility of feedback, accountability to use feedback, social awareness through feedback, and self-efficacy in dealing with feedback (Linderbaum & Levy, 2010). The argument is that just as organizations or departments can set or create different FEs, individuals can differ in their receptivity to or ideas about feedback. Together, FE and FO seem important in understanding and affecting the feedback process. FO is related to other individual differences such as *learning goal orientation,* which is the extent to which

feedback orientation (FO)
An individual's overall receptivity to feedback.

individuals believe learning and development are possible. In addition, employees high on FO also tend to seek feedback more often (Linderbaum & Levy, 2010) and perform better than those low on FO (Dahling, Chau, & O'Malley, 2012). In fact, a recent study found each dimension of the FO to be positively related to satisfaction with the feedback (Rasheed, Khan, Rasheed, & Munir, 2015). In a study of 623 Chinese employees, researchers examined how age might relate to various dimensions of FO and impact feedback reactions (Wang, Burlacu, Truxillo, James, & Yao, 2015). They found that older workers had higher levels of the social awareness dimension and lower levels of the utility dimension of FO than did younger workers. Further, dimensions of the FE related to feedback reactions in ways that were driven by age and FO dimensions. In sum, FO differences across age resulted in different patterns of the relationships between feedback characteristics and feedback reactions (Wang et al., 2015).

Perhaps in the most interesting study of all those that link FE with FO, researchers found that the experience of a favorable or encouraging FE is not a positive thing for all people. In a study of correctional facility employees, they found that for those individuals who were low on FO, a positive FE was not a motivating factor, and in some instances it actually resulted in a negative effect on motivation (Gabriel et al., 2014). In other words, if you are someone who does not value feedback and is not very receptive to it, then working in an environment that values feedback and encourages feedback seeking may be a toxic combination for you and is likely to result in a lack of motivation and a good deal of dissatisfaction. This is the only study to examine and find this interesting effect. More research is certainly needed.

With recent research on FE and FO, we are gathering more and more advanced knowledge about the feedback process. This promises to improve both the employee development and performance appraisal processes in organizations.

Legal Issues in Performance Appraisal

It is illegal in the United States to discriminate in performance appraisals on the basis of non–performance-related factors such as age, gender, race, ethnicity, religion, and disability. In short, no employee can be promoted, demoted, fired, transferred, or laid off on the basis of any of these factors.

In their article on legal requirements and technical guidelines related to performance appraisal, James Austin, Peter Villanova, and Hugh Hindman (1995) make the following recommendations for performance appraisal:

1. Start with a job analysis. Austin and his colleagues especially recommend developing criteria from a documented job analysis.
2. Communicate performance standards to employees in writing.
3. Instead of relying on just an overall rating of performance, recognize that there are separate dimensions of performance that should be evaluated on an individual basis.
4. Use both objective criteria and subjective judgments where possible, and ensure that the subjective judgments evaluate job-related behaviors rather than global personality traits.

5. Give employees access to an appeal mechanism.
6. Use multiple raters rather than one rater with absolute authority.
7. Document everything pertinent to personnel decisions. This recommendation can't be emphasized enough, as one never knows who is likely to bring suit and on what basis. The courts have been clear in such cases: If there is no documentation to support the personnel decision in question, the defendant (i.e., the organization that is being sued for discrimination) is likely to lose the case.
8. Where possible, train the raters; if training is not possible, provide them with written instructions for conducting the performance appraisal.

In an extensive review of court decisions between 1980 and 1995, Jon Werner and Mark Bolino (1997) uncovered more than 300 court of appeals decisions related to performance appraisal. Of these, 109 were related to age discrimination, 102 to race discrimination, and 50 to sex discrimination. The authors attempted to determine which factors were most closely associated with judgments made for or against the plaintiffs (i.e., the employees bringing suit against the company). What they concluded was that, in general, the practical recommendations provided in the literature (e.g., Austin et al., 1995; Barrett & Kernan, 1988; Malos, 1998) seemed consistent with the ways in which the courts make these judgments. For instance, judgment was more likely to be made for the defendant when (1) a job analysis was used, (2) written instructions regarding the appraisal process were provided, (3) employees were provided with an opportunity to review their appraisals, (4) multiple raters agreed on performance ratings, and (5) rater training was used. In addition, these researchers found that rating format (e.g., BARS versus graphic rating scales) didn't seem to matter, as this topic never came up in the judges' written decisions.

On the other hand, the court's view of the fairness of the appraisal process was extremely important. Folger, Konovsky, and Cropanzano (1992) recommend the use of the *due process metaphor* view of performance appraisal, in which the emphasis is on (1) adequate notice, (2) a fair hearing, and (3) judgments based on evidence. Each of these three components of due process—a familiar term in the legal arena—is built around perceptions of justice. In short, courts value the presence of justice in appraisal systems—an observation that supports continued research on contextual variables of this type. A recent examination of the role of due process in performance appraisal reviews 20 years of work in this area that shows the value of this approach (Levy, Cavanaugh, Frantz, & Borden, 2015).

An interesting trend that promises to get the attention of legal experts working both in and out of organizations is the use of wearable technology, or what has become simply known as "wearables." As I write this chapter, I can look down at my arm and see a Fitbit on my wrist that records my steps, my intense activity, my sleep, my workouts, and so on. My iPhone is next to me on my desk and I can use it to pull up the data recorded by my Fitbit—for instance, it tells me that I slept 5 hours and 36 minutes last night, that it took me 15 minutes to fall asleep, and that I woke up once and was restless for about 21 minutes (clearly, I was worried about finishing this chapter by today's deadline!). This technology has been available for years and has certainly caught on in the healthcare market as a way for individuals to monitor their behaviors. It also provides a way for employers to offer incentives for specific healthy behaviors (e.g., my wife gets a rebate on her health insurance if she hits certain annual goals for healthy behaviors). In recent years companies have become interested in using wearables for workplace purposes, such as increasing productivity, enhancing employee well-being, and reducing work-related injuries (Rackspace, 2014). In a recent study conducted by the University of London and online hosting company Rackspace, the findings clearly show that wearables (e.g., brain activity sensors, motion monitors, and posture coaches) can benefit both productivity and well-being at work. Theatro, one of the leading designers of wearables, reports a 5% to 10% increase in productivity for companies that have used their wearable device (Dishman, 2014).

Wearables have become a $20 billion industry (Harrop, Hayward, Das, & Holland, 2015) and is expected to grow to $27 billion by 2022 (Lamkin, 2018)! Although the most common use is currently in the health arena, other workplaces are finding ways to use wearable technology. For instance, in manufacturing industries, wearables can be used for hands-free access to information such as instructions, maintenance guides, real-time interactions with offsite supervisors or trainers, and so on (Openshaw & Greenspun, 2014). H. James Wilson (2013), a leading researcher on human–technology interaction, has coined the term *physiolytics* to describe the link between wearables and data analysis to provide feedback and improve performance. He suggests three kinds of analysis that come from wearables: (1) quantification of movements within physical work environments, such as monitoring and recording goods or products at a distribution center, (2) working with information more efficiently by analyzing the time and motion spent on a task, and (3) analyzing our personal "big data," such as blood pressure, brain waves, and cognitive patterns. A recent case study using Theatro technology found that training costs were reduced by $1.3 million through the use of in-ear technology, which delivers training content directly to trainees on the sales floor. Compared to traditional computer-based training, these trainees were more engaged and able to learn in an environment that resembled the one in which they will work. Theatro itself claims that $3.3 million of increased sales resulted from the repurposed labor resulting from keeping the employees on the sales floor (Kanithi, 2018).

There is a great deal of agreement that the technology of physiolytics has potential for individuals and organizations, but as you can imagine, there are questions of privacy and worries that "Big Brother" will be watching our every move. Privacy concerns are likely to lead to legal issues and lawsuits filed by employees claiming that their rights have been violated (McLellan, 2015; Tremaine, 2015). The use of wearables for organizational purposes is still in the early phases, and the jury is out on how well this technology will be received by employees, but I think there is very little doubt that cases will begin making their way through our court systems. It will be very interesting to watch what emerges from this legal activity, considering industry experts argue that by 2022 the total sales of wearables will surpass 233 million units with some of this coming with the increase of smart clothing and earwear (Lamkin, 2018).

Summary

I/O psychologists often play an active and important role in performance management systems, which include performance appraisal. Performance appraisals are used to make personnel decisions, to provide employees with important job-related feedback, and to document employee performance as a way of protecting the organization from potential legal suits.

Performance appraisals are available in many formats, including graphic rating scales, BARS, CARS, checklists, and employee comparison procedures. Each format has advantages and disadvantages (see Table 5.1); none has clearly been identified as the single best approach. The process of appraising an individual's performance is very complex and can result in rating errors such as halo, leniency, central tendency, and severity. Two types of rater training, Rater Error Training (RET) and Frame-of-Reference (FOR) training, have been established as potentially useful in reducing these errors and/or improving accuracy.

Contemporary performance appraisal research continues to focus on the social-psychological context in which the appraisal takes place. This research indicates that ratee and rater reactions toward the appraisal process are important in measuring the success of an appraisal system. Also relevant to the appraisal process are the relationship between the supervisor and subordinate, the political climate within the organization, the rater's motivation and accountability, and the degree of trust among raters.

Participation is another integral element in the appraisal process. One interesting development in this area is the 360-degree feedback system, in which employees receive performance feedback from peers, subordinates, supervisors, and clients/customers. Self-ratings are commonly included in this system, especially when the only other participant is the supervisor. Although 360-degree feedback is relatively new, research indicates not only that many employees like it but also that it has a great deal of potential for improving the feedback process in general. Employee development continues to grow in importance as both employees and organizations recognize the benefits of employee growth and learning. Employee development not only benefits individual workers but also functions as a source of competitive advantage for organizations that encourage and provide opportunities for development.

Adherence to legal guidelines is also critical to the performance appraisal process. Recent work has identified fairness as important not only to ratees but also to the courts, which appear to weight it quite heavily in making judgments for or against plaintiffs. And while emerging technology, such as wearables, will lead to many interesting and innovative approaches to performance management, it may also raise legal concerns over privacy violations.

As we have seen, performance appraisal has widespread implications for organizations because appraisal information is used for such important personnel decisions as promotions, demotions, layoffs, dismissals, and raises. Obviously, invalid performance appraisal information is likely to result in poor organizational decisions. Our understanding of the appraisal process is thus enhanced by our appreciation of the context in which it takes place.

Key Terms

BARS (p. 131)
central tendency (p. 141)
coaching (p. 125)
context (p. 146)
critical incidents (p. 132)
distributional errors (p. 140)
feedback environment (FE) (p. 155)
feedback orientation (FO) (p. 155)
Frame-of-Reference (FOR) training (p. 144)
halo (p. 139)
leniency (p. 140)
performance appraisal (p. 125)
performance management (p. 125)
Rater Error Training (RET) (p. 143)
severity (p. 141)
telework (p. 130)
360-degree feedback (p. 127)
true halo (p. 140)
upward appraisal ratings (p. 128)

TAKING IT TO THE FIELD

Although this chapter primarily discusses rating formats, errors, and training, performance appraisal also depends heavily on appropriately defining the criteria/criterion, as discussed in Chapter 4. Jia Xiang, who is the director of Camp Bear Paw, a youth summer camp, is seeking advice on how to evaluate her camp counselors. Please read her e-mail and then respond to her questions, carefully considering both the content and the format of her measure.

> Thank you for being willing to lend your expertise on this project. I had a meeting with the leadership and the counselors to come up with a list of things we believe are important qualities of counselor performance:
>
> - Watchful: Look out for problems so they can be prevented rather than addressed after they occur.
> - Self-motivated: Don't need to be told what to do and proactively seek out ways to help.
> - Fun: We find that if someone is willing to be creative and engaging they do well.
> - Firm: Need to be willing to say no and to keep kids in line if they are doing something dangerous or inappropriate.
> - Fair: Don't play favorites with the kids, and keep disagreements with fellow counselors professional and private.
>
> Can you draft a 10-item performance appraisal to use with our counselors? I'd also be interested in your thoughts about what rating scale we should use. In the past, we've used a 1 (very poor) to 5 (excellent) rating scale, but nearly every counselor gets a 4 or a 5, so it hasn't been very useful.
>
> Thank you!
> Jia Xiang

Critical Thinking Questions

1. Based on what you have read, design a rater training program that will help supervisors assess the performance of their subordinates as accurately as possible. What techniques will you include? What important variables can impact rater effectiveness?
2. The chapter discusses both performance appraisal and the broader performance management system. How do these two concepts differ? How are they related?
3. Several different rating formats are presented in the chapter. If you were designing a performance appraisal process for an organization, what types of scales or formats might you use (e.g., BARS, graphic scales)? Would you use paper-and-pencil formats, or would you use a technology-based approach? What rationale do you have for your choices?
4. Multisource feedback is a tool used to obtain feedback from individuals at various levels of the organization. Frequently, the target employee (the one receiving the feedback) selects the individuals who will provide feedback. What are some potential drawbacks of allowing the target employee to select his or her raters? Should the target employee's supervisor be responsible for selecting raters? How else could raters be determined? How might we incorporate technology into this process?
5. Consider a time you received feedback from someone on your performance. Was it helpful? What did they do (or what could they have done) to help you understand how to use their feedback to improve?
6. Consider a job you have had in the past, or a job you would like to have. How might you use wearables to measure your performance in this role? How would you feel about that company monitoring your behaviors using this technology?

Application Questions

1. Imagine that when using a 360-degree performance appraisal system, an employee receives a rating of 3.5/5 from his supervisor, 4.2/5 from his peers, and 2.25/5 from his subordinates. What might be some reasons for these discrepancies? What would you recommend that this individual's supervisor do to gain a better understanding of what these ratings indicate about the employee's performance?
2. Imagine that, in order to assign grades on a group project, your professor requires all group members to provide ratings for one another, which he will consider in grading the project. One group member has been particularly troublesome—she has been argumentative, refuses to accommodate other group members' schedules, and completes the bare minimum amount of work. Would you feel comfortable giving her a poor performance appraisal? Why or why not?

3. Once again, imagine that your professor has asked for your feedback on your fellow group members' performance on a project. This time, you have a group member who is very proactive in volunteering and takes an active role; however, even though he tries very hard, he has a lot of trouble understanding concepts introduced in the class; as a result, often the work he does is not usable for the project. Would you feel comfortable giving him a poor performance appraisal? Why or why not?
4. In many work environments, employees can engage in organizational citizenship behaviors (acts that help coworkers or the organization, even though they are not part of their job), such as staying late to work on a project they are not normally a part of or speaking positively about the organization to other people. Should these behaviors be considered in a performance appraisal? Why or why not?
5. Think of a time when you worked with someone who did not perform as well as you believed he or she should. What sorts of things did this person do or not do that led to a poor performance? If you were asked to deliver feedback to this person, what would you say?

Suggested Readings

Allen, T. D., Golden, T. D., & Shockley, K. M. (2015). How effective is telecommuting? Assessing the status of our scientific findings. *Psychological Science in the Public Interest*, *16*(2), 40–68. A nice review of the telecommuting trend and one that comes at it from a scientific perspective.

DeNisi, A. S., & Murphy, K. R. (2017). Performance appraisal and performance management: 100 years of progress? *Journal of Applied Psychology*, *102*(3), 421. An exhaustive and detailed review of the performance management literature and a look to the future.

Fletcher, C., and Baldry, C. (2015). Multi-source feedback systems: A research perspective. In I. T. Robertson and Cary L. Cooper (Eds.), *Personnel psychology and human resources management: A reader for students and practitioners.* New York: Wiley. A very nice review of the empirical research on 360-degree feedback and the implications of this work, written in a way that makes the material easy to grasp.

Levy, P. E., Tseng, S. T., Rosen, C. C., & Lueke, S. B. (2017). Performance management: A marriage between practice and science—just say "I do." In *Research in personnel and human resources management* (pp. 155–213). Bingley, West Yorkshire: Emerald Publishing Limited. This recent piece summarizes major concerns of practitioners about performance appraisal, links these concerns to the existing literature, and recommends organizational changes and future research.

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management, 30*(6), 881–905. This comprehensive review of the performance appraisal research since 1995 places a particular emphasis on the social context of performance appraisal.

Schleicher, D. J., Baumann, H. M., Sullivan, D. W., Levy, P. E., Hargrove, D. C., & Barros-Rivera, B. A. (2018). Putting the system into performance management systems: A review and agenda for performance management research. *Journal of Management*, *44*(6), 2209–2245. This is the most thorough review of performance management out there and it provides a systems model that helps conceptualize how the important elements fit together.