

4 THE FOUR KEY CHALLENGES TO PERFORMANCE APPRAISAL

Learning Objectives

- 4.1 Learn about the challenges supervisors and managers face when trying to evaluate the performance of their subordinates
- 4.2 Understand the ways both proximal and distal context factors influence the way performance appraisal is conducted and used in organizations
- 4.3 Understand the range of uses of performance appraisal and the potential for these uses to come into conflict
- 4.4 Consider the factors that motivate supervisors and managers to inflate the performance ratings they provide

Performance appraisals have a very high failure rate; up to 90% of appraisal systems in organizations are viewed as ineffective (Pulakos, Mueller-Hanson, Arad, & Moye, 2015; Smith, Hornsby, & Shirmeyer, 1996). Performance management systems do not seem to fare much better; the conclusion that performance management is broken is shared among many researchers (Pulakos & O’Leary, 2011; Pulakos et al., 2015). Given the very high rate of apparent failure of performance appraisal and performance management systems, it is unlikely that the shortcomings of these systems are simple or easy to fix. Rather, this sort of failure rate suggests that there are broad, systematic problems with the way we typically evaluate job performance, provide feedback, and act on that feedback. In this chapter, we describe in a general way the four most important reasons why performance appraisal systems fail in organizations. This chapter lays out broad conceptual arguments about why performance appraisal systems fail. In [Chapters 5–12](#), we examine in more detail the research, theory, and insights from practical applications of performance appraisal that are relevant to understanding these four seeds of failure. In [Chapters 13 and 14](#), we discuss ways of creating performance appraisal systems that might overcome at least some of these challenges, and that provide organizations an opportunity to obtain useful measures of the performance of their employees.

There are several obvious shortcomings in the performance appraisal systems in many organizations. First, they are often expensive and time-consuming. For example, Buckingham and Goodall (2015) tallied the amount of time spent at Deloitte on performance appraisal (e.g., meetings, time spent filling out forms) and concluded that over *2 million hours* per year were spent evaluating the performance of 65,000 employees. This investment of time and resources might be justified if managers and employees saw concrete benefits, but in many organizations very little actually happens as a result of performance appraisals. Thus, performance appraisal often looks like a lot of pain for very little gain. Employees do not feel that they receive useful feedback, and when they *are* given feedback, they do not often change. No matter how much

time and effort is invested in performance appraisal, there is little clear evidence that performance appraisal is actually contributing to the success of individuals or organizations.

We see issues such as the cost of performance appraisals and the lack of clear benefits to individuals and organizations as symptoms or indicators of the failure of performance appraisal, not as root causes. That is, high costs and questionable benefits are *symptoms* of failure, but they do not explain *why* appraisal systems appear to fail in so many organizations. In our view, the problem is not with the surface features of appraisal systems, such as the types of scales used to evaluate performance, the way raters are trained, or the frequency of appraisal and feedback meetings. Rather, there are four structural causes that we believe are at the heart of the failure of performance appraisal systems.

The Structural Causes of Failure

Our review of research, theory, and practice in the area of performance appraisal leads us to conclude that there are four core problems at the heart of most performance appraisal systems in organizations. First, the task of evaluating performance, even under the most favorable conditions, is a very difficult one. Second, there are numerous layers of contextual variables, ranging from broad national cultures to the climate and culture of the organization, to the nature of the relationships between individual raters and ratees that complicate the design, implementation, and interpretation of performance appraisals, and that allow nonperformance factors to systematically influence the evaluation of individual performance. Third, organizations often use the same performance appraisal system for multiple purposes, and these purposes are often in conflict. Finally, there are a number of social and political factors in organizations that motivate most raters to give high ratings, regardless of the rater's judgments about the performance and effectiveness of each ratee. For example, 99.9% of the federal employees examined in a recent Government Accounting Office report were rated "Fully Effective" or higher (GAO, 2016). This pervasive rating inflation undermines many of the potential uses of performance appraisals, and efforts to overcome this inflation (e.g., by relying on ranking or forced distributions rather than on performance ratings) create their own problems in the measurement of job performance.

The Task of Measuring Performance Is Inherently Difficult

Suppose you supervise seven skilled production workers in a small electronics manufacturing facility. You are asked to evaluate the performance of these seven employees each year. Doing a good job of measuring performance entails several steps, each of which can be challenging. First, you must obtain information about each of your subordinates and have some level of assurance that this information is representative of their overall performance. It is common for supervisors to have limited opportunities to directly observe their subordinates as they perform their job, especially if workers are dispersed in ways that makes it impossible to observe all employees simultaneously. You might try to solve this problem by walking through the workplace on an irregular schedule (this is one component of what Peters and Waterman, 2004, describe as "management by walking around"), but it is likely that your subordinates will know (at least part of the time) when they are being watched, and that they will be on their best behavior when you

are in sight. Thus, while you might obtain a lot of information by walking around and observing, you will always have to worry about the possibility that what you see is not representative of what happens on a regular basis in the workplace.

Having obtained information, you must do two things: (1) make sense of what you have seen (encoding information), and (2) recall at some later time what you observed. Performance appraisals are often an annual affair (often performed in conjunction with the determination of pay raises), so you might have to encode and recall a lot of information. We know a great deal about human memory ([Chapter 5](#) reviews this research), and it is likely to be a significant challenge to correctly recall information over a much shorter time, much less recalling information about performance that occurs over the course of a year.

Suppose you clear the first several hurdles and are able to obtain, make sense of, and accurately recall representative information about each subordinate's performance. You will next be faced with the problem of integrating all of that information. Most employees will exhibit a mix of good and poor performance over time, perhaps showing distinct strengths and weaknesses, or perhaps simply showing some inconsistency in their behavior. Your task will be to integrate information about performance that may vary over time, and that may exemplify a mix of things an employee typically does very well and things he or she typically does less well.

The next challenge supervisors face is to apply consistent and correct evaluative standards. It will not do you much good if you correctly observe, encode, recall, and integrate information if your definition of what constitutes good versus poor performance is both idiosyncratic and inconsistent with the standards other members of the organization rely upon.

Of course, obtaining good measures of performance is only the start of what it would take to make performance appraisal effective. DeNisi and Pritchard (2006) presented a detailed model that relates features of performance appraisal and performance management systems with the motivation to improve performance. This model suggests that the links are far from simple. For example, they note the following:

- Performance appraisal and performance management systems that make the relationship between an employee's actions and the results of those actions clear will be associated with higher levels of performance improvement.
- Performance appraisal and performance management systems that increase the employee's capability to perform well (e.g., through a developmental focus) or that provide adequate resources and authority to employees to allow them to perform their jobs will be associated with higher levels of performance improvement.
- Performance appraisal and performance management systems that result in high levels of agreement across rating sources will be associated with higher levels of performance improvement.
- Performance appraisal and performance management systems that provide clear expectations and performance standards will be associated with higher levels of performance improvement.
- Performance appraisal and performance management systems viewed as fair and unbiased will be associated with higher levels of performance improvement.

- Performance appraisal and performance management systems that provide consistent evaluations over time will be associated with higher levels of performance improvement.
- Performance appraisal and performance management systems that consistently link performance with rewards and sanctions will be associated with higher levels of performance improvement.
- Performance appraisal and performance management systems that produce outcomes that are consistent with employees' needs and preferences will be associated with higher levels of performance improvement.

As [Chapters 5–12](#) show, there are barriers to virtually all of the suggestions listed above, and it is difficult to develop a performance appraisal or performance management system that provides good measures of performance *and* makes appropriate use of those measures.

Thus, the first of the seeds of failure for performance appraisal systems is that they ask supervisors or other raters to successfully execute a very challenging task—to arrive at well-founded and consistent evaluations of the performance of each subordinate, and then turn these evaluations into useful performance feedback. Even under optimal conditions (e.g., where there are few stressors or distractions and where the task of observing and evaluation performance is a top priority), it is difficult to do this consistently and well. Under more normal conditions, where performance evaluation is only one of many tasks the supervisor is charged with and where there are multiple distractions and barriers to successful evaluation, this task is even more challenging. Most supervisors, managers, and executives do not work under conditions that are especially favorable for completing the task of performance appraisal, and it is no wonder that the measures that *are* obtained are somewhat suspect.

Finally, it is worth remembering that the task of performance appraisal is not only difficult in a technical sense, it is also a stressful and emotionally draining task (Pearce & Porter, 1986). In several later chapters, we will discuss in some detail research on the discomfort raters and ratees experience when they confront the task of performance appraisal, and especially the task of giving and receiving feedback. Performance appraisal is one of the least favorite activities for raters, and it is hardly a picnic for the employees they rate.

Contextual Factors Influence Appraisals

Performance appraisal does not occur in a vacuum, but rather in the context of many important and impactful variables, ranging from the values of the society in which the organization is located to the climate, culture, and norms of the organization, to the particular values, assumptions, and experiences of the supervisor and his or her subordinates. All of these can influence the way performance appraisal is carried out and used in organizations, as well as the reactions of managers, supervisors, and employees to the performance appraisal system. A number of authors have emphasized the importance of contextual variables (e.g., Judge & Ferris, 1993; Mitchell, 1983; Murphy & Cleveland, 1995), and research on contextual influences on performance appraisal and performance has emerged as an important area of study. This research is reviewed in detail in the next chapter.

The reason why contextual factors are important is that they influence the way people implement, use, and interpret performance appraisal. Thus, the performance appraisal systems in two different organizations might both yield equally accurate and relevant information about the performance of the people being rated, but the way that information is interpreted, used, or ignored might vary considerably because of differences in national cultures, organizational policies and practices, or even supervisor–subordinate relationships.

The term “context” can refer to variables at a number of different levels of analysis, as illustrated in [Figure 4.1](#). Most generally, there are two categories of contextual influences: (1) distal context variables that have a generally indirect influence on performance appraisal, and (2) proximal variables that are more directly related to the performance appraisal practices and to the effects of performance appraisal. For example, both broad cultural variables (e.g., the extent to which a culture is collectivistic versus individualistic) and legal and economic variables (e.g., whether personnel practices are legally regulated, whether the job market is very tight or very loose) can indirectly influence performance appraisals by helping to set conditions that make different approaches to conducting and using performance appraisal more or less likely to succeed. More proximal variables, such as the organization’s policies regarding how and why it conducts appraisals, or the nature of the relationship between supervisors and subordinates in a particular work group, have a more direct influence on the way performance appraisal is conducted and used and the reactions of users to the performance appraisal system. Thus, the same appraisal system, placed in different contexts, may end up looking, feeling, and working in quite different ways.

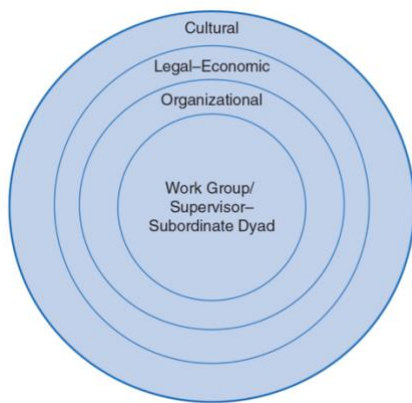


Figure 4.1 Levels of Contextual Variables

Two of the most widely studied facets of contexts or environments are turbulence and munificence (Katz & Kahn, 1978; March & Simon, 1958; Thompson, 1967). Turbulence refers to the extent to which the environment changes or remains stable. For example, Katz and Kahn (1978) note that the economic environment of Western nations was highly turbulent in the 1970s, with unpredictable swings between inflation and depression. Turbulent environments are thought to lead to uncertainty in organizations; this uncertainty could affect performance appraisal practices. For example, organizations facing uncertainty may be less willing to commit resources to developing employees and may place less emphasis on developmental feedback in their performance appraisal systems. Munificence refers to the degree of scarcity versus abundance of

critical resources in the environment. For example, when unemployment is high, the labor market is, from the organization's point of view, a munificent environment. When the organization is flush with money, the environment for creating performance-based rewards is munificent. Scarcity of critical resources is likely to lead to conflict within organizations (March & Simon, 1958). The implications of this conflict will depend largely on which environments are munificent, and which are plagued by scarcity.

For example, if the labor market is very tight, performance appraisals may be used quite differently than if the physical environment is characterized by abundance. If the labor market is very tight, organizations will have an incentive to reward and retain good performers, whereas a looser labor market might make organizations more open to terminating poor performers and attempting to replace them with better-qualified job applicants. Katz and Kahn (1978) noted five aspects of the environment that organizations must monitor and respond to in order to be effective: (a) societal values, (b) the political/legal environment, (c) the economic/labor environment, (d) the information/technological environment, and (e) the physical/geographical environment. Each of these aspects of the environment is likely to affect performance appraisal in some way, with some having a more direct impact on appraisal (i.e., legal environment may act like a proximal cue, that directly influences rater behavior), and others (e.g., societal values) having a very indirect, but nevertheless important, effect on performance appraisals.

How Does the Context Affect Appraisal?

Murphy and Cleveland (1995) identified six variables that appeared most likely to mediate the relationship between contextual variable (particularly distal context) and performance appraisal: (1) performance standards, (2) the choice of performance dimensions, (3) the frequency of appraisal, (4) the quality of supervisor–subordinate relations, (5) the consequences of receiving high versus low ratings, and (6) the perceived legitimacy of the appraisal system. These six variables do not exhaust the list of possible mediating variables, but they do give us a basis for describing the different ways in which various aspects of the environment might indirectly affect performance appraisal.

Environments are complex, and different aspects of the environment could have different effects on each of the intervening variables listed above. To help structure our discussion of environmental effects, we will restrict our attention to the effects of the munificence of the five aspects of the environment identified by Katz and Kahn (1978). Munificence can be defined in the following ways for each aspect of the environment: (a) societal—the extent to which socio-political norms and values support the concept of performance appraisal, as typically practiced in organizations; (b) legal—the extent to which the legal system facilitates and allows typical performance appraisal practices; (c) economic—extent to which general economic conditions are favorable for the organization; (d) technical—extent to which an organization possesses or controls the technology and work methods needed to carry out the organization's functions; and (e) physical—the extent to which necessary physical resources are available.

Murphy and Cleveland's (1995) review suggested that some contextual variables had stronger effects on particular mediators than on others; some of their more important conclusions are summarized in [Table 4.1](#). For example, the question of whether we should expect all workers to

perform well, or should we set standards that allow some workers to succeed and others to fail might depend on the values of society as a whole. A society that emphasizes competition and achievement might lead to different (probably higher) performance standards than one that emphasizes satisfaction with work. The legal system probably also affects performance standards, although to a lesser degree. That is, the legal system may not have a direct effect on standards (i.e., raising or lowering standards), but it is likely to affect the degree to which standards are made explicit, as well as the degree to which they are enforced. A legal system that is basically antagonistic to performance appraisal will probably lead to the formulation of explicit, concrete standards that are communicated directly to employees.

We would expect different (higher) performance standards during a business recession than during a period of economic growth or stability. We would also expect the technological position of a company to affect standards. A company that is behind the times technologically might try to compensate by setting high standards; a company that is technologically sophisticated might not see strong links between the performance standards enforced for individual workers and the output of the organization. Companies that employ assembly-line technologies might have different performance standards than companies that rely on semi-autonomous work teams to produce their product. Finally, the effects of physical munificence are likely to be very similar in strength to those of economic munificence, but perhaps in a different direction. When the physical resources needed for production are not available, it may be necessary to adjust performance standards downward (Peters, O'Connor, & Eulberg, 1985).

The definition of what constitutes performance and the relative importance of different parts of the performance domain are likely to be affected by the environment. A comparison of American versus Japanese management styles provides an example of societal influences. An American manager is likely to emphasize individual task performance, whereas a Japanese manager is likely to emphasize ability to work with the group when evaluating the performance of the same subordinate (Pascale & Athos, 1981). The legal system will probably also have a moderate influence. As the legal system becomes less munificent (i.e., less favorable to current appraisal practices), we predict that organizations will emphasize dimensions that involve clear, concrete, objective outcomes, and will de-emphasize interpersonally oriented performance dimensions.

Table 4.1 Effects of Environmental Munificence on Three Aspects of Appraisal Systems

Aspects of Appraisal	Aspects of Environment				
	Societal	Legal	Economic	Technical	Physical
Standards	Strong	Moderate	Strong	Strong	Strong
Dimensions	Moderate	Moderate	Strong	Strong	Strong
Consequences	Moderate	Strong	Strong	Weak	Moderate

The munificence of the economic environment will strongly affect the choice of performance dimensions. In an unfavorable economic environment, we expect that organizations will define performance in terms of short-term, bottom-line goals. Thus, a supervisor who is not good at producing products but who is good at developing subordinates might receive different evaluations, depending on the economic environment. The technical environment will also have a strong impact on the choice of performance dimensions. The importance of technically oriented

dimensions will probably vary across industries, and across companies with varying levels of technical sophistication.

The munificence of the physical environment will strongly affect the extent to which performance is defined in terms of efficiency versus output. When resources are scarce, the worker who produces less but conserves materials and resources might receive a more favorable evaluation than a worker who produces a great deal but is wasteful. The reverse might be the case when resources are cheap and plentiful.

One of the key determinants of whether the supervisor gives high or low ratings is the consequences that are attached to the ratings. Rating inflation is quite likely when ratings have a strong influence on salaries, promotions, and so on, and is least likely if ratings are not used to make administrative decisions (see [Chapters 11](#) and [12](#)). All five of aspects of the environment will have at least a moderate effect on the consequences of giving high versus low ratings; for two of the five aspects, this influence may be substantial.

Spotlight 4.1 Failure to WARN

A Unique Performance Appraisal Lawsuit

Gregory Anderson, a former employee of Yahoo, filed a suit alleging that Yahoo manipulated its performance appraisal system and that he received lower-than-deserved performance ratings, leading to his separation from the company. This lawsuit illustrates some of the unexpected legal conflicts that can arise as a result of the way organizations use performance appraisal.

In 2102, Yahoo adopted a performance appraisal system that required managers to evaluate the performance of each employee, comparing him or her to immediate peers.¹ Employees were placed in one of five categories (e.g., Greatly Exceeds, Occasionally Misses, Misses); the proportion placed in each category is (the lawsuit claims) fixed in advance by the organization. Employees in the lowest category are liable to be terminated.

While the facts of each case always differ, there is nothing unusual about a lawsuit alleging that a performance appraisal system has treated some employee or group of employees unfairly. What makes this case unique is that it claims that Yahoo used this appraisal system to sidestep the federal Work Adjustment and Retraining (WARN) Act. This law requires companies to give advance notice of reductions in force and to give the employees who will be laid off or released certain benefits and opportunities for alternative employment. The suit claims that Yahoo shed over 1,000 employees between late 2014 and early 2015, claiming that these employees were released due to poor performance, and that this policy resulted in a de facto layoff. In essence, by forcing a certain number of employees into the lowest performance category each year, Yahoo might be able to lay off as many employees as it wanted without triggering the WARN Act by pretending that employees were let go for performance reasons.

Employment litigation often focuses on claims of discrimination (this layoff also includes such claims), but the legal context of performance appraisal is more complex, sometimes including issues that are covered by a range of state and federal laws that have little to do with employment

discrimination. Human resource managers need to be aware of the broader legal environment when designing performance appraisal systems.

¹ <http://www.lexology.com/library/detail.aspx?g=024a6698-f4f7-4300-b195-363ad858a49a>

Societal norms and values will have some influences on the consequences of high versus low ratings; the extent of this influence will be determined by the degree to which societal values support or discourage treating good versus poor workers differently. For example, the question of whether individuals should be rewarded in proportion to their level of performance would probably be answered differently in the United States than in Japan.

The legal system will strongly affect the degree to which performance appraisals are likely to be connected to administrative decisions. A legal system that is hostile to performance appraisal, as it is typically practiced, will also be hostile to the use of performance ratings as a primary basis for allocating organizational rewards. A legal system that strongly acknowledges management's rights to evaluate the performance of workers will probably also support linking rewards directly to performance ratings.

The lower the munificence of the economic environment (and, to a lesser extent, the physical environment), the greater the incentive to tie rewards to performance levels. We would predict that organizations will be more likely to use performance appraisals to make administrative decisions, ranging from salary and promotions to layoffs, when the economic environment is unfavorable than when it is favorable. In part this is because organizations have less slack (i.e., less ability to survive given substandard performance) in an unfavorable economic environment than in a favorable one.

In the next chapter, we will examine the influence of contextual variables, as well as the results of research on contextual influences on performance appraisal in some detail. For now, simply understanding that contextual variables influence the design, implementation, interpretation, and use of performance appraisal systems is enough to help you understand why these systems so often fail. The same performance appraisal system may work well in some contexts and poorly in others, making it hard to identify best practices that will actually work across the many contexts in which performance appraisal systems are found.

The Conflicting Purposes of Performance Appraisal in Organizations

Ratings are used for many purposes in organizations (e.g., to determine promotions and salary increases, to evaluate developmental needs, to serve as criteria for validating tests and assessments, to provide documentation for legal purposes), and these purposes often come into conflict (Cleveland, Murphy, & Williams, 1989; Murphy & Cleveland, 1995). Fifty years ago, Meyer, Kay, and French (1965) noted the incompatibility of using performance appraisal simultaneously to make decisions about rewards and to provide useful feedback, and suggested that the different uses of performance appraisal should be separated by creating different evaluative mechanisms for rewards versus feedback. The rewards versus feedback distinction continues to be an important one, but there are many uses of performance appraisal in organizations that have little to do with rewards or feedback.

Cleveland et al. (1989) reviewed studies published in several scientific and professional journals, and identified 20 different uses for performance appraisals in organizations; these uses are shown in [Table 4.2](#). They surveyed organizations to determine the relative importance of these 20 uses for performance appraisal. Of the 20, the most important included using performance appraisals for salary administration, for identifying individual strengths and weaknesses, and for documenting personnel decisions. More to the point, every organization surveyed indicated that performance appraisals were used for multiple purposes.

Table 4.2 Twenty Uses of Performance Appraisal

Salary administration
Promotion
Retention and termination
Recognition of individual performance
Layoffs
Identify poor performance
Identify individual performance
Performance feedback
Determine transfers and assignments
Identify individual strengths and weaknesses
Personnel planning
Determine organizational training needs
Evaluate goal achievement
Assist in goal attainment
Evaluate personnel systems
Reinforce authority structure
Identify organizational development needs
Criteria for validation research
Document personnel decisions
Meet legal requirements

Cleveland et al. (1989) noted that two of the critical uses of performance appraisal focus on using performance ratings to either identify differences *between* individuals (i.e., determining which ratees perform better than others, overall) or differences *within* individuals (i.e., determining individual strengths and weaknesses), and that many organizations use performance appraisals for both purposes. This is potentially problematic, because these two purposes can be in conflict. In particular, just about everything you do to enhance accuracy and discriminability when identifying differences between people will tend to detract for accuracy and discriminability when identifying individual strengths and weaknesses. To understand why, consider the two sets of performance profiles illustrated in [Figure 4.2](#). The top set of profiles illustrates performance patterns that make it easy to distinguishing between employees; employee 1 is clearly the better performer. In this set of profiles, each employee is at a pretty

similar level on all four performance dimensions, and if one employee is pretty good at most aspects of his or her job and another is not so good, it will be relatively easy to distinguish between them. The bottom set of profiles shows performance patterns for employees who have strong strengths and weaknesses. Because of the peaks and valleys in the performance profiles, the overall performance levels of these two employees will end up being highly similar; they will both be “average” performers, even though their strengths and weaknesses are so different. A system that highlights relative strengths and weaknesses tends to produce a lot of overall ratings that are “average,” making it hard to distinguish one employee from another in terms of their overall performance level.

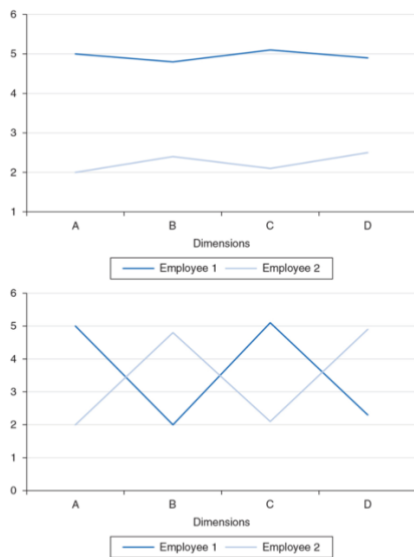


Figure 4.2 Performance Profiles Illustrating Between- Versus Within-Person Differences

The implications are clear. If you work hard to identify individual strengths and weaknesses, you will end up with performance profiles that classify many employees as pretty similar in terms of their overall performance level. If you work hard to identify who is the best and who is the worst performer, and where everyone else fits in that performance distribution, you will tend to end up with relatively flat profiles, which are great for distinguishing between people but almost useless for identifying individual strengths and weaknesses.

Figure 4.2 Performance Profiles Illustrating Between- Versus Within-Person Differences

The implications are clear. If you work hard to identify individual strengths and weaknesses, you will end up with performance profiles that classify many employees as pretty similar in terms of their overall performance level. If you work hard to identify who is the best and who is the worst performer, and where everyone else fits in that performance distribution, you will tend to end up with relatively flat profiles, which are great for distinguishing between people but almost useless for identifying individual strengths and weaknesses.

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

other uses (e.g., the range restriction that is likely when most ratees receive high ratings might limit the value of appraisals as criteria for validating selection tests). One of the general conclusions of Cleveland et al. (1989) is that if an organization tries to do too many things with its performance appraisal system, it might end up doing *all* of them poorly.

There are many reasons why it is common for the same performance appraisal system to serve many purposes, but one is that it is hard to tell who “owns” performance appraisal. At one level, this is a system that is likely to be developed and maintained by the human resources department. At another level, it is an important part of the interaction between individual supervisors and their direct reports. At yet another level, it is a system that has potentially diagnostic value across the entire organization. For example, suppose you find that the average rating on a 5-point scale (where “4” indicates that the employee exceeds expectations and “5” indicates that the employee far exceeds expectations) is a 4.25 and that 80% of all employees receive an overall rating of 4 or 5 (this is not an unusual outcome). There are three possibilities. First, maybe the employees of this organization *are* that good. Second, it is possible that expectations are simply set too low. Third, and most likely, there is rampant rating inflation; inflation appears to be so common that you will need an organization-level explanation for this phenomenon (i.e., the explanation must involve forces that cross boundaries within the organization). This type of rating inflation is in fact quite common, and it forms the fourth explanation for the frequent failure of performance appraisal and performance management systems.

Raters Are Motivated to Distort Performance Appraisals

Imagine once again that you are the supervisor of seven employees, and it is your task to rate their performance. Based on your observations and the information you have collected, you conclude that one employee exceeds expectations for his or her job, five meet expectations, and one falls a bit short of expectations. Should you give these three groups ratings of 4, 3, and 2, respectively? The answer depends on two things. First, what goal or goals are you pursuing when you complete the performance appraisal form? If your goal is to provide accurate ratings, it is pretty clear what you should do (i.e., you should turn in one 4, five 3s, and one 2). On the other hand, suppose your goal is to motivate your employees to perform better next year. You might conclude that you are more likely to achieve this goal if you give high ratings than if you give low ones, *even when low ratings are deserved*.

Think about the possible consequences if you give a poor-performing employee the 2 he or she deserves, versus the likely consequences if he or she gets a higher rating, and perhaps a better raise or better protection from an upcoming layoff. If this employee gets a low rating, it is a good bet that he or she will resent this and see it as unfair; [Chapters 9](#) and [11](#) review research that will show why this is such a common reaction. This employee will probably get a lower raise and may feel more vulnerable. It is unlikely *any* of these will motivate this employee to improve, and it is likely that these will all contribute to an antagonistic relationship between you and this employee. On the other hand, if you give this employee an inflated rating, a raise is more likely, there is less likelihood that the rating will be seen as biased or unfair, and the employee will be more secure and will possibly even feel some gratitude to you. It might be quite reasonable for a rater to purposefully give inflated ratings, because the consequences of inflated ratings might be

much more positive (and perhaps even better for the organization) than the consequences of accurate ratings.

Performance appraisal is a tool for influencing the behavior of the individuals being evaluated, and a supervisor would be a fool to not at least *consider* the possibility that giving inflated ratings might be a more effective management strategy than giving accurate ratings. Put another way, if you are a manager, what would you choose: (1) a strategy that is likely to lead to resentment, lower motivation, and worse relations with your subordinate (accurate ratings), or (2) a strategy that is likely to improve your relationship with subordinates and to motivate them to perform better (rating inflation)? We do not see this as a difficult choice.

Of course, while giving inflated ratings might improve some outcomes that *you* care about as a manager, giving accurate ratings might improve some outcomes that are important to the organization, such as providing accurate criteria to be used when validating selection tests or assessments or when justifying decisions that are made on the basis of ratings. There are probably some uses of ratings where inflated ratings are better than accurate ones and others where the reverse is true, and there are likely to be some instances where top management will want raters to act against their self-interest to provide accurate rather than inflated ratings. Whether and under what conditions this expectation might be a reasonable one will be examined in [Chapters 13](#) and [14](#).

Summary

If performance appraisal failed in some organizations and was seen as a rousing success in others, we might try to identify specific aspects of the appraisal system that caused it to fail, such as bad rating scales or inadequate training. By all indications, the failure of performance appraisal and performance management systems is the norm rather than the exception, and where there is broad systematic failure, it is best to look for broad systematic causes. Our review of more than a century of performance appraisal research suggests that there are at least four broad causes for the failure of performance appraisal systems.

First, under even the most favorable circumstances, it is hard to accurately evaluate the performance of several individuals over time. Performance rating is rarely done under favorable circumstances; rather, the observation and evaluation of performance is just one of many duties most supervisors and managers take on. Second, context matters, so that an appraisal system that seems successful in one setting might fail dismally in another, and the same intervention (e.g., more rater training) might have quite different effects in different settings. Context is a complex construct that can be defined at many levels, and the effects of some context variables (e.g., national cultures) might work in the opposite direction of the effects of others (e.g., organizational culture).

Third, performance appraisal systems are used for many purposes in organizations, and these purposes may place conflicting demands on the system. In this chapter, we have highlighted the conflict between two uses of performance appraisal—that is, for distinguishing *between* individuals versus distinguishing *within* individuals, but as we will show

in [Chapter 8](#), there are many possible purposes and many possible sources of conflict surrounding performance appraisal.

Finally, the dynamics of performance appraisal in many organizations set up a number of incentives to inflate ratings, while at the same time creating disincentives for accuracy in performance appraisal. These incentives sometimes involve the rater's self-interest (e.g., a rater who claims that all of his subordinates are performing well also ends up looking like a successful manager), but organizations can benefit from rating inflating as well, at least if inflation is more likely than accuracy to help supervisors maintain positive relations with their subordinates and to motivate employees to perform better in the future. More than 30 years ago, Banks and Murphy (1985) noted that raters do not turn in inflated ratings because they *can't* accurately evaluate performance, but rather because they *do not want to*. In several of the chapters that follow, we will examine research on what motivates raters to provide accurate versus distorted ratings, and how this research can be used to help build better performance evaluation systems.

Exercise: Maintain a Behavior Diary

Performance appraisal researchers have tried a number of strategies for making the task of observing and evaluating the behavior of employees more manageable. One that has received consistent support is the use of behavior diaries. Several studies have shown that taking notes about the behaviors different employees show in the workplace can make evaluation more reliable and valid. However, the task of maintaining a behavior diary is not a simple one. The best way to get an appreciation of the costs and benefits of behavior diaries is to try this method out.

For a week, try and take reasonably frequent notes about the behaviors of four or five individuals who you regularly observe. The notes should be in behavioral form, describing *what* each person did. For example, if you are observing call center workers, something like “Jane presented several potential solutions to the caller she was working on” would be a behavioral note, whereas “Jane had a very good call” would not be behavioral information. At the end of a week, ask yourself several questions:

- What did you write down?—The best behavior diaries will include a mix of behaviors, some illustrating good, poor, or average performance incidents, but many behavior diaries are dominated by incidents of very good or very bad performance.
- How much did you write down?—You might find yourself starting out by taking notes on a regular basis, but also find out that the notes start to tail off quickly.
- Are your observations representative?—Do you record behaviors pretty much as they occur, or only the outstanding and unusual ones?
- Are your observations driven by prior impressions?—If Joe is generally a good performer and Sally is generally a poor performer, you might find that you only record good behaviors for Joe and poor performance instances for Sally.
- How intrusive is the task?—Time and energy you devote to maintaining a behavior diary might have been spent doing other things. What sorts of tasks are displaced by the need to record behaviors?

We think a short “test drive” will give you a concrete sense of the costs and benefits of maintaining behavior diaries. While the research on the potential contributions of this method is both solid and supportive, many of the studies of this method involve students whose only task is to take diaries, and who do this over a fairly short period. Before you commit yourself or your organization to this strategy, we think it is very valuable to try the method out and make an informed judgment about whether or not this strategy is likely to be helpful in your organization