7 HOW CONTEXT INFLUENCES PERFORMANCE APPRAISAL

Learning Objectives

- 7.1 Learn how national cultures and legal and political environments influence the way performance appraisal is conducted and used
- 7.2 Understand how the work environment influences the supervisors' ability to observe and evaluate the performance of their subordinates
- 7.3 Learn how organizational climates influence performance appraisals
- 7.4 Understand the way supervisor–subordinate relationships influence and are influenced by performance appraisals
- 7.5 Understand why performance appraisal policies and practices that work in one organization may not work as well in another

The way performance appraisal is designed, implemented, used, and interpreted is strongly influenced by the social and organizational context in which performance appraisal occurs. Imaging two organizations, one that is seen as highly successful, that has lots of resources and that operates in a nation whose culture strongly values individual excellence, and the other that is in an authoritarian country with a failing economy and that provides poor pay and worse opportunities. Even if all of their performance appraisal practices (e.g., rating scales, rater training, frequency of appraisal, individual feedback meetings with the supervisor) are the same, it is safe to assume that these two appraisal systems will function quite differently, will be used for different purposes, and will be interpreted by organizational members very differently. In this chapter, we explore the different levels of context and the contextual influences that are most likely to have an impact on performance appraisal in organizations.

We start by defining contexts, noting that there are both aspects of the broad society in which performance appraisal operates (distal context) and aspects of the organization itself (proximal context) that influence the design, implementation, and use of performance appraisal systems. We examine two distal factors in some detail, national cultures and the legal environment within which organizations operate. We next examine a number of proximal factors, ranging from the climate and culture of the organization to the structure and nature of work groups to the structure of performance appraisal systems (e.g., who provides input, how often is performance evaluated). One broad theme of this chapter is that the success or failure of performance appraisal systems often depends as much upon these contextual factors as it does on the design and implementation of the appraisal system itself.

The Emergence of Context-Oriented Research

Landy and Farr's (1980) review of performance appraisal research published between 1920 and 1980 helped ignite a steady stream of research on cognitive processes in performance appraisal. This work proved to be extremely useful for understanding how judgments about performance are made and how difficult it could be to accurately perceive, remember, and make sense of performance over the long periods of time typical for performance appraisals (usually done yearly). By the early 1990s, the weaknesses of this cognitive research were becoming apparent. First, this work had little relevance to the problems faced by organizations that were trying to improve their appraisals (Banks & Murphy, 1995; Murphy & Cleveland, 1991). More important, this research paid scant attention to the organizational context within which appraisals are conducted in organizations (Bretz, Milkovich, & Read, 1992; Ilgen, Barnes-Farrell, & McKellin, 1993), but rather focused on some basic principles of judgment that might cut across organizations or even across nations. In the last 30–40 years, considerably more attention has been paid to the influence of contextual variables on performance

appraisal (Arvey & Murphy, 1998; Bernardin & Beatty, 1984; DeCotiis & Petit, 1978; Ferris, Munyon, Basik, & Buckley, 2008; Landy & Farr, 1980; Levy & Williams, 2004; Murphy & Cleveland, 1991, 1995).

Aycan (2005) points out paying attention to the context is by no means unique to performance appraisal research; a wide variety of theories of organization and management outside of the domain of performance appraisal have long been concerned with the way context shapes the behavior of organizations. For example, *general systems theory* (Katz & Kahn, 1978) was developed to help explain the way the effects of organizational policies and practices are influenced by the social, legal, and political contexts in which organizations operate. *Strategic fit* models (e.g., Schuler & Jackson, 1987) help to explain how human resource policies and practices interact with the broader strategy of the firm. *Agency theory* (Jensen & Meckling, 1976) examines the way the titles and descriptions of jobs and occupations (e.g., managerial versus non-managerial job titles) influence HR policies and practices. Kanungo's *model of culture fit* (Kanungo & Jaeger, 1990) was developed explicitly to examine the interplay between societal cultures and internal work cultures in determining HR policies and practices and their effectiveness. Thus, the trend in performance appraisal research to emphasize context can be thought of as a specific instance of a much broader trend in the organizational sciences over the last 50 years of paying careful attention to the way the context in which organizational policies are carried out influence the implementation, interpretation, and effects of those policies.

What Is Context?

One of the definitions of the term "context" that we find most useful is that the context is the "setting within which something exists or happens, and that can help explain it."[1] This definition has two appealing features. First, as applied to performance appraisal, context is the setting or the set of circumstances that surrounds the practice of performance appraisal in organizations. Thus, context is "bigger" than the thing it surrounds, but that thing (here, performance appraisal) cannot be fully separated from all that surrounds it. That is, performance appraisal cannot be taken out of its normal context without changing important aspects of the appraisal process, and research that ignores this context may not tell us much about how performance appraisal actually works in organizations. Second, context has explanatory power. That is, performance appraisal is designed, implemented, used, and interpreted in particular ways in part because of the context in which it occurs. The more we know about important aspects of this context, the better we are likely to understand performance appraisal.

Table 7.1 Multiple Levels of Context

Distal	National culture
	Legal and economic system
	Economic environment
	Technical environment
	Physical environment
Proximal	Organizational climate and culture
	Organizational strategy
	Work group characteristics
	Supervisor–subordinate relationships

The context in which performance appraisal exists can be defined at many different levels of analysis. It is common in performance appraisal research (see, for example, Murphy & Cleveland, 1995; Murphy & DeNisi, 2008) to distinguish between distal and proximal contexts, where the proximal context refers to variables within the organization while the distal context refers to variables that exist outside of the organization. Distal context usually does not directly affect the practice or the use of performance appraisal, but it sets conditions that will strongly influence this process (Murphy & Cleveland, 1995). Proximal context, on the other hand, has a more direct effect and may play a direct role in forming or influencing the use and interpretation of performance appraisal in organizations. Table 7.1 lists distal and proximal influences on performance appraisal that have been studied most extensively in the last 30 years. Each of these is examined in the sections that follow.

Distal Context

The first of the contextual variables listed in Table 7.1, national culture, has been the focus of a considerable amount of research, some of which deals with performance appraisal and some of which deals with the more general question of how cultures should be defined and measured. Considerable progress has been made in measuring national cultures and in developing and testing theories of the fit between human resource practices and national culture.

National Culture

The most useful model for studying national cultures is based on the widely cited work of Geert Hofstede (See Hofstede, 1980; Hofstede, Hofstede, & Minkov, 2010; See House et al., 2004 and Schwartz, 1999 for similar models). On the basis of his analysis of over 100,000 questionnaires completed by employees of IBM in 72 countries, Hofstede (1980) argued that national differences in work-related values and preferences could be understood in terms of a small number of dimensions: (1) Power Distance—social inequality, or the degree of separation between people at different levels in organizations or groups; (2) Individualism–Collectivism—orientation toward individuals versus groups; (3) Masculinity–Femininity—preference for achievement and assertiveness versus cooperation and caring; and (4) Uncertainty Avoidance—tolerance for versus avoidance of ambiguity and lack of control. His later work added the dimensions of: (5) Long-Term versus Short-Term Normative Orientation—focus on immediate versus more distant outcomes, and (6) Indulgence versus Restraint—orientation toward free expression and satisfaction of desires versus suppression of these desires (Hofstede, 2001; Hofstede, Neuijen, Ohayv, & Sanders, 1990), but the relevance of the first four dimensions for performance appraisal has been studied more extensively.

Power Distance[a]		Uncertainty Avoidance	
100		100	
80	Arab Countries	80	Japan
	France		
60		60	Germany
40	USA, Netherlands	40	USA
			Great Britain
20		20	

Individualism		Masculinity	
100		100	
	USA		Japan
80		80	
	France		
60		60	USA
40	Brazil	40	France
20	Hong Kong	20	
	Indonesia		Netherlands

Long-Term Orientation		Indulgence	
100	Hong Kong	100	
80	Japan	80	New Zealand
			USA
60	Brazil	60	Belgium
40	Netherlands	40	
20		20	China
	West Africa		

[a] Approximate values on all dimensions from Hofstede (1994), Hofstede, Hofstede, & Minkov (2010).

Figure 7.1 Exemplars of Hofstede Dimensions

Hofstede and his colleagues have measured these factors in several large-scale studies and have collected an extensive database that allows them to describe national cultures for over 75 nations. Figure 7.1 gives some examples of countries that exhibit high, medium, and low levels of the six Hofstede factors.

The potential relevance of the Hofstede culture dimensions to performance appraisal is clear. Consider, for example, some of the decisions that might be involved in designing a performance appraisal system, such as who should evaluate performance, how often, with what tools, and for what purpose. Each of these is likely to be influenced by one or more of the dimensions of national culture Hofstede identified. Table 7.2 identifies dimensions of national culture that might be most relevant to particular aspects of performance appraisal.

Table 7.2 Dimensions of National Culture Potentially Relevant to Design of a Performance Appraisal System

Design Feature	Relevant Dimension	Rationale
Who Provides Ratings	Power Distance	Self-ratings and upward feedback are unlikely to be acceptable if there is a large distance between supervisors and subordinates
How Often	Uncertainty Avoidance	Systems with more structure and predictability (e.g., annual appraisals are likely to be preferable to more informal systems)
What Type of Rating Scales	Power Distance	Methods that involve goal setting may be inconsistent with strict hierarchies because they place responsibility for defining performance on the ratee
	Uncertainty Avoidance	Discomfort with ambiguity may make detailed behavioral scales more acceptable
Purpose of Rating	Individualism	Preference for individual versus group evaluations and for using performance appraisal as a basis for promotion and salary may vary as a function of individualism–collectivism

For example, in a culture that is high on Power Distance, there is considerable respect for hierarchy and for understanding one's place in the scheme of things. Most crucially, in societies where Power Distance is high, the low-power members of that society accept the existing hierarchy as natural and reasonable. In these societies, power relationships are paternalistic, autocratic, and centralized, and bosses are expected to make unilateral decisions that will be accepted and respected by all members of the organization. In a nation where Power Distance is high, top-down appraisals will seem sensible and natural, whereas appraisals that are participative will simply not fit, and will not be accepted by supervisors or subordinates.

A culture that is high on Collectivism emphasizes identification with the group (e.g., the family, work groups) as well as a high degree of loyalty to that group and a strong aversion to things that separate the individual from the group. Team-based performance appraisals are much more likely to be accepted in a Collectivistic culture than in an Individualistic culture (which places more emphasis on the individual), whereas programs to recognize individual achievements may not be accepted as sensible in a Collectivist society.

A culture that scores high on Uncertainty Avoidance is likely to value strict laws, policies that are uniformly applied, and predictability, and performance appraisal practices that are regular and well-defined (e.g., annual appraisals and feedback meetings) are likely to feel more comfortable and sensible to members of this culture than evaluation practices that are unstructured and unpredictable (e.g., the practice of giving organizational members frequent but irregular feedback as part of a performance management program). Rating scales that spell out what performance dimensions and performance levels mean (e.g., behaviorally anchored rating scales) are likely to be more acceptable than more abstract scales that leave much to the judgment of the individual rater.

Cross-National Performance Appraisal Research

There have been numerous studies reporting the characteristics of performance appraisal systems in different countries (see, for example, Amba-Rao, Petrick, Gupta, & Von Der Embse, 2000; Arthur, Woehr, Akande, & Strong, 1995; Grund & Sliwka, 2009). While some of the specific features of performance appraisal systems (and other human resource systems) common in particular countries are a function of differences in economic and legal systems, cross-national differences in performance appraisal practices and outcomes appear to be largely a function of differences in national cultures (Aycan, 2005). For example, Maley (2013) notes that the purpose of performance appraisal is likely to vary across nationalities and cultures; several studies discussed below provide support for this prediction.

Human resource management practices of 35 countries in the Organization for Economic Development and Cooperation (OECD) are summarized at http://www.oecd.org/gov/pem/hrpractices.htm. OECD data confirm there are differences in the extent to which performance appraisals are done and in the ways in which these appraisals are used. For example, performance appraisals are used frequently in the United States, but even more frequently in Japan and Denmark. Performance appraisals are used to make promotion and salary increase decisions in the United States, but the use of appraisal as a tool for making decisions about salary is less common in Brazil, and is quite rare in Russia. These cross-national differences are likely to be influenced by differences in the economic and legal systems of the countries involved, but cross-national cultural differences appear to be most important.

As noted earlier, the purpose of appraisal varies across countries (Maley, 2013). A study examining 10 different countries regarding their performance appraisal purposes revealed that, for example Canada, Latin America, Taiwan, and the United States emphasize performance appraisal for pay decisions more than Australia, Indonesia, and Japan (Milliman, Nason, Zhu, & De Cieri, 2002). Chiang and Birch (2010) found that communication-development purposes were most frequent in Finland and Sweden, and formal feedback was more prevalent in low collectivist countries, such as the United Kingdom, the United States, and Canada.

Aycan's (2005) review of research of human resource practices led to some specific propositions regarding performance appraisal, some of which have received considerable support in subsequent research. For example, Aycan (2005) proposed the following:

- In collectivist, high Power Distance cultures, interpersonal competencies (e.g., teamwork facilitation, respect for others) and process (e.g., intentions, motivation) will be more important than task related competencies and processes.
- In low Power Distance cultures, performance will be regular and formalized and involve multiple sources, but high Power Distance cultures will be less structured and will depend more on the opinions and perceptions of supervisors and superiors.
- In collectivist cultures, group feedback will be preferred over individual feedback, and feedback will be indirect and non-confrontational.

Tests of National Culture Hypothesis

Discussions of and theories about the roles of national cultures in shaping human resource practices are common, but rigorous empirical examinations of these effects are less common. Peretz and Fried's (2012) study is a noteworthy exception. These authors examined the effects of national culture on performance appraisal, absenteeism, and turnover in 21 countries. In particular, they examined the relationships between power distance, individualism–collectivism, uncertainty avoidance, and future orientation (Hofstede, 1980) on appraisal system features and on some of the potential outcomes of performance appraisal. Table 7.3 summarizes some of their key findings. For example, both Individualism and Power Distance are related to

a preference for a single rating source as opposed to multi-rater systems, such as those frequently used to provide 360-degree feedback.[2]

Table 7.3 Empirically Supported Relationships Between Cultural Variables and Performance Appraisal System Features

Higher Power Distance	Lower likelihood of using multiple sources (e.g., peers, supervisors)
Higher Individualism	Lower likelihood of using multiple sources
Higher Uncertainty Avoidance	Higher likelihood of having a formal PA system
Higher Future Orientation	Higher likelihood of using multiple sources Larger proportion of workforce is evaluated Higher likelihood of having a formal PA system

A related study by Adsit, London, Crom, and Jones (1997) suggested that the acceptability of upward feedback varies across countries, and that this variance can be accounted for (at least in part) by national differences in power distance, individualism–collectivism, uncertainty avoidance, and masculinity. Similarly, Chiang and Birch (2010) presented data suggesting that the purpose of appraisal (e.g., salary, promotion, identify strengths and weaknesses) and the formality of appraisal systems varies across nations, and that some of this variability can be accounted for by the cultural values identified by Hofstede (1980). There is evidence that the acceptability of appraisal systems that involve upward feedback or even lateral feedback (e.g., peer ratings) varies as a function of culture (Entrekin & Chung, 2001).

A number of papers have examined the effects of presumed East–West cultural differences on performance appraisal practices and outcomes. This research highlights the danger in making broad assumptions about East–West differences, for example that members of Asian cultures are collectivist, whereas members of American and European cultures are individualistic. For example, Snape, Thompson, Yan, and Redman (1998) compared performance appraisal practices and attitudes in Britain and Hong Kong. Their results suggest that British performance appraisals tended to be more participative and more focused on development, whereas Hong Kong appraisals tend to be more authoritarian, with a stronger focus on evaluation and performance-based rewards. They also found that employees in Hong Kong support this reward-oriented application of performance appraisal, an attitude that seems inconsistent with the assumption that collectivism reigns across all of Asia. Similarly, Yu and Murphy (1993) showed that the assumption that Asian employees will be modest in their self-ratings of performance (as opposed to the usual pattern of inflated self-appraisals in U.S. organizations) is not always a sound one, and that collectivism-inspired modesty may be unique to specific countries or cultures within Asia.

Claus and Briscoe's (2009) review suggests that despite the assumption that broad cultural differences will lead to large differences in the way human resource systems such as appraisal are implemented and used, the data tend to indicate more convergence than differentiation. For example, Shadur, Rodwell, and Bamber (1995) compared HR practices in Japanese versus non-Japanese corporations operating in Australia. In the 1990s, it was widely believed that the unique cultural features of Japanese organizations gave them a genuine competitive advantage. While this might be true (or might have been true at the time; Japan is no longer seen as the type of cultural and economic juggernaut it once seemed to be), whatever cultural differences existed, they did not translate into large differences in human resource practices. While cultural values are likely to matter, they may have a great impact on attitudes toward human resource practices than on the practices themselves.

Culture and Multisource Rating Systems

Evaluation plays a critical role in defining power and hierarchy in organizations; the defining characteristic of a supervisor is that he or she has the right and responsibility to evaluate the performance of subordinates.

Multisource rating systems, which collect information about and evaluations of performance from supervisors, peers, customers, and even sometimes subordinates, violate this hierarchical principle, allowing people who do not have formal authority over a ratee to participate in his or her evaluation. It is likely that this sort of inversion of traditional hierarchy is more acceptable in some cultures than in others, especially cultures that differ in terms of power distance or collectivism–individualism. Furthermore, it is likely that the characteristics of ratings from different sources, in particular the mean and the intercorrelations among ratings, are influenced by broad cultural variables. For example, in high Power Distance and in high Collectivism cultures, subordinate ratings tend to be both higher and more intercorrelated than peer or supervisor ratings (Ng, Koh, Ang, Kennedy, & Chan, 2011).

Power Distance and Collectivism can sometimes have opposite effects on the use of and reactions to multisource rating systems. For example, there are countries that are high on both Power Distance and Collectivism (e.g., Indonesia[3]); the use of peer ratings would be discouraged if Power Distance is high but encouraged if Collectivism is high. Because these two facets of culture lead to different conclusions about the acceptability of multisource systems, it is difficult to make confident predictions about how such systems might be received in this country.

Cross-Cultural Differences in Multinational Organizations

Multinational organizations are becoming an increasingly important part of the world economy, and these organizations face a number of challenges in designing human resource systems (Maley, 2013). The challenge in designing sustainable and effective performance appraisal systems is especially complex for multinational organizations whose various divisions are characterized by different cultures (DeNisi & Smith, 2014; Kostova, 1999; Murphy & DeNisi, 2008). Cross-cultural differences in individualism and collectivism appear to be particularly important in determining whether specific human resource management practices will work across borders (Erez, 2011).

Multinational organizations face competing pressures in deciding whether to adopt consistent human resource management practices across their various divisions or to develop unique systems for each. On the one hand, there is considerable efficiency and potentially better control of the costs and outcomes of different human resource systems if similar systems can be used across national boundaries (Schmid & Kretschmer, 2010). On the other hand, human resource systems that work well in some nations might be culturally inappropriate in others. To add another layer of complexity, multinational organizations differ in the extent to which they tend to impose uniform human resource practices across their divisions in different countries versus adapting their human resource systems to local cultures and norms. Comparisons of European-owned versus American-owned multinational corporations suggest that American-owned multinationals are more likely than European-owned multinationals to tolerate diversity in the human resource systems used in their international affiliates (Cleveland, Gunnigle, Hearty, Morley, & Murphy, 2000; Gunnigle, Murphy, Cleveland, Hearty, & Morley, 2002). However, the use of human resource practices that are based on the expectations of the home country are likely to be employed in evaluating top managers (Janssens, 1994).

The finding that cultural variables can influence performance appraisal system design and use should not be taken to indicate that every country will follow its own unique practices. For example, Claus and Briscoe (2009) note broad across-nation similarities (e.g., performance appraisal systems in Ghana and Nigeria are more similar than dissimilar to those in the United States; see also Arthur et al., 1995) and within-nation heterogeneity (they note differences within India; see also Amba-Rao et al., 2000). On the other hand, similarities or differences in cultural factors do not always predict differences in human resource management practices. Even within regions of the world that are broadly similar in terms of general culture factors such as individualism–collectivism, there can be significant variation in the way human resource practices such as performance appraisal are carried out and used (Paik, Vance, & Stage, 2000).

The increase in multinational organizations and the increasing frequency of expatriate assignments implies that there will be an increasing number of rater–ratee pairs where raters and ratees are of different nationalities (Caliguiri & Day, 2000). First, raters and ratees may have different expectations regarding the nature of performance (e.g., the relative importance of task and contextual performance) and in the specific behaviors that are expected under each of the major dimensions of performance. Caliguiri and Day (2000) suggest that ratees who are higher on self-monitoring are more likely to make appropriate adjustments as they move into cultures where expectations regarding performance may vary as a function of nationality.

As multinational corporations become more dominant, the need to develop performance appraisal systems that serve the needs of the organization and its members and that make sense in the cultures where they are applied is becoming increasingly apparent. One particularly difficult issue is developing appraisal systems for expatriates. As more and more managers undertake international assignments, the question of how to evaluate their performance has become more pressing. Martin and Bartol (2003) suggest that appraisal systems for expatriates are most likely to be successful if the clarify expectations, if appraisals are frequent, if they are seen as fair, and if they have a developmental focus.

Limits on the Effects of Broad Cultural Variables

Mishra and Roch (2013) note that while there is important variability in cultural values across nations and regions of the world, there is also often considerable variability within nations, which implies that even in the same nation, some raters are likely to be higher or lower on particular dimensions of culture (e.g., power distance) than others. In particular, they argue that within national cultures, there are often differences in the extent to which raters adopt an independent versus an interdependent self-construal, and that this difference might influence performance appraisals. Raters who view themselves as independent tend to view their own behavior, and by extension the behavior of others, as being caused by individual differences in abilities, skills, motivation, and the like. Raters who view themselves and others as interdependent tend to emphasize the role of situational factors and constraints in explaining behavior. These differences translate into predictable differences in rating behavior and in feedback; raters who take an interdependent view tend to give higher ratings and tend to avoid giving feedback that focuses on differences in performance.

Although there is clear evidence that cultural variables have an impact on the way performance appraisals are designed, conducted, and used, it is not so clear that culture has a strong effect on performance ratings themselves. Ployhart, Weichmann, Schmitt, Sacco, and Rogg (2003) studied the measurement equivalence of performance rating from three countries with substantially different cultures, Canada, South Korea, and Spain. They found similar factor structures and general evidence of measurement equivalence, although there was more error variance in ratings of several performance dimensions in some countries than in others. Nevertheless, these data suggest more communality than distinctness, and that performance ratings might be more comparable across cultures than has sometimes been thought.

Even in cases where there are broad cultural effects on performance ratings, these effects are often not as simple and straightforward as they seem. For example, collectivism is typically associated with leniency in performance ratings, but Barron and Sackett (2008) note that it is important to distinguish between in-group and institutional collectivism. Institutional collectivism focuses on societal and organizational practices that encourage and reward collective action and the collective distribution of rewards, and institutional collectivism is correlated with leniency in rating. In-group collectivism, on the other hand, focuses on pride, loyalty, and cohesiveness in organizations and families, and this type of collectivism does not have clear implications for leniency in rating.

Legal and Political Environment

Passage of the Civil Rights Act of 1964 dramatically changed the legal environment of performance appraisals in the United States. Prior to 1964, it was difficult for any employee to legally contest his or her performance evaluation, no matter how subjective, capricious, or inaccurate the appraisal. The reason for this is that there was no legal requirement that private organizations accurately evaluate workers' performance. The same is true today, except in the case where a performance measure has adverse impact on some protected group (typically minorities or women). Adverse impact occurs when a test, measurement device, or procedure has systematically different outcomes for different groups. For example, if most men passed a strength test, but most women failed, the test would have adverse impact on women if it was used to make hiring, firing, promotional decisions, and so on. Similarly, if you found that across many jobs men received systematically different performance ratings than women, it would constitute evidence of adverse impact. Any measure that has, or that might have, adverse impact is potentially vulnerable to legal scrutiny, and if adverse impact is demonstrated, the employer must establish the validity or business necessity of that measures, or in the case of performance appraisal, of the performance appraisal system. That is, the law allows companies to use tests, assessments, or appraisal systems that result in systematically different scores for member of different demographic groups (e.g., old versus young workers, men versus women, members of different racial and ethnic groups), only if they demonstrate that the scores on these measures are related to differences in performance or effectiveness on the job. For example, if you use a cognitive ability test in selecting among job applicants, it is possible that the test might lead to different selection rates for different racial and ethnic groups, but if this test can be shown to be job related, the law allows you to use this test.

Spotlight 7.1 Performance Appraisal in Japan

In this chapter, we have discussed the role of broad cultural and legal systems in shaping the way performance appraisal is conducted and used. To illustrate this phenomenon more concretely, it is useful to examine some similarities and differences in the performance appraisal systems used by firms in the United States and Japan. Shibata (2000, 2002) describes some of the distinctive features of Japanese performance appraisal systems.

Like many American organizations, Japanese organizations often conduct annual appraisals of their employees. However, because pay in blue-collar positions is determined by a combination of skill levels and age[4], ratings tend to focus on broad abilities, skills, and attitudes (e.g., diligence, cooperation). Work attitudes of white-collar employees are not assessed in a comparable way, perhaps on the assumption that most white-collar employees exhibit uniformly high levels of the required work attitudes.

Endo (1998) notes an interesting difference between American and Japanese performance appraisal systems— that is, that Japanese employees are not always informed about the ratings they receive. In American organizations, employees almost always know the performance ratings they receive, a design feature that may lead to higher levels of rating leniency (Shibata, 2002). It is possible that the relatively high levels of Power Distance in Japan makes it easier to conduct evaluations without informing employees of the results of those evaluations, but other cultural factors may push performance appraisal systems in more egalitarian directions. For example, Shibata (2002) notes that the appraisal systems used for blue-collar and white-collar workers tend to be more similar in Japanese than in American organizations. Furthermore, it is unlikely that blue-collar employees in Japanese firms who receive the lowest ratings will be laid off (as oppose to the "rank-and-yank" systems championed by Welch and Byrne, 2001) or that high-rated employees will be put on a fast track for promotion.

In the 1930s, many Japanese firms incorporated American performance appraisal practices (Endo, 1998), but the appraisal systems in these two countries have evolved in different directions. Shibata (2002) notes that the very difficult economic situation confronting Japan from the 1990s has probably accelerated changes in Japanese performance appraisal systems, perhaps making them less distinctively Japanese and more closely targeted toward market forces than toward the national culture. Nevertheless, it seems clear that differences in

economic systems (e.g., different levels of unionization) and cultural values can have a strong and lasting impact on the way performance appraisal is administered and used.

Legal issues connected with performance appraisal are reviewed by Bernardin and Beatty (1984), Cascio (1987), Cascio and Bernardin (1981), and Latham (1986).[5] Some general guidelines for complying with federal anti-discrimination laws in carrying out performance appraisals are listed in Table 7.4. All of these guidelines can be thought of as safeguards designed to protect employees from arbitrary or subjective evaluations that may more closely reflect the rater's biases than they reflect the ratee's performance.

Table 7.4 General Guidelines for Legal Compliance in Performance Appraisal

1. Performance appraisals should be based on specific dimensions whose relevance has been established through job analysis.
2. Raters should receive training or instruction.
3. Performance dimensions should be defined in terms of behaviors.
4. There should be feedback to the ratee, and an appeal process for ratings the individual feels are inaccurate.
5. Raters should have adequate opportunities to observe the performance they will be asked to evaluate.
6. Extreme ratings should be documented.
7. If possible, there should be multiple raters.
8. Appraisals should be frequent—at least annually.
9. Employees should have meaningful input into their evaluations (e.g., self-ratings).

Sources: From Bernardin & Beatty (1984); Feild & Holley (1982); Werner & Bolino (1997).

There is clear evidence that the legal environment affects the way in which performance appraisals are done. For example, Cleveland, Murphy, and Williams (1989) surveyed over 100 corporations to determine the degree to which performance appraisal was used for various purposes in organizations. In their confirmatory factor analysis of 20 potential uses, one of the four factors is labeled "Documentation," and refers to the uses of performance appraisals to meet federal, state, and local requirements, in particular to justify adverse personnel actions (e.g. firing a worker for poor performance). Their surveys showed that documentation represented one of the most frequent uses of performance appraisal. Although Cleveland et al. (1989) did not investigate the issue, we suspect that the use of performance appraisal to document personnel decisions is highest in organizations that are willing to fire poor performers, and is lowest in situations where it is not possible to take action in the face of performance that is at least marginally acceptable (e.g., where a strong union prevails, or where contractual provisions make for-cause firing difficult). Only the most foolish employer will now dismiss a worker for poor performance without first collecting clear and convincing documentation of poor performance, as well as documentation of efforts on the part of the organization to provide opportunities for and encouragement of improvements in performance.

There have been numerous court cases in which performance appraisal figured meaningfully either as a personnel practice that was being challenged or as a criterion used to evaluate other personnel practices. In general, the same features that are thought to influence the perceived fairness of performance appraisals (e.g., giving employees a voice in appraisal, giving raters clear instructions and training, review of results with ratees, triangulation among multiple raters) appear to influence the outcomes of litigation (Werner & Bolino, 1997). The bottom line for organizations is that it is in their best interests to make sure that the performance appraisal systems they employ are clearly job related (in the sense that there is evidence that they are designed to carefully measure performance on the job, as opposed to yielding global judgments about individuals that cannot be reliably linked to the job itself) and fair.

Most discussion of legal issues in performance appraisal focus on the United States, which makes sense given the size and importance of the U.S. market and the uniquely American reliance on litigation as a method of law enforcement. However, many nations and groups of nations have their own laws and policies dealing with human resource practices such as performance appraisal. For example, the directives of the European Union

provide some guidance to individual members establishing minimum standards for labor law (e.g., equal pay for men and women is mandated by EU directives), but at the same time European organizations operate under a number of different legal and regulatory regimes.[6] Sisson and Storey (2000) note that a number of countries follow the "Roman-Germanic system," which prioritizes statutory regulation, whereas other members follow a "Nordic system," based on the emphasis of collaborative agreements characteristic of Denmark and Sweden. There has been movement within the EU to develop a more uniform pattern that draws from both systems. However, the different political and legal systems in Nordic nations versus most other European nations may tend to encourage more participatory systems in countries like Denmark or Sweden than in, for example, France or Italy.

Countries that follow different political and legal systems often differ in other ways that makes it difficult to isolate the effects of political and legal considerations. In particular, political and legal differences are often confounded with cultural differences. For example, several papers describe performance appraisal in China (e.g., Cheng & Cascio, 2009; McEvoy & Cascio, 1990), and these often make the point that different aspects of Chinese culture lead to specific effects on the use and interpretation of performance appraisal in Chinese companies. For example, Chinese companies appear to be less likely than their non-Chinese counterparts to have traditional top-down performance appraisal systems and to be more likely to use multiple sources of input and to include both on-the-job and off-the-job behavior as components of performance appraisal. These features of performance appraisal in Chinese companies tend to be attributed to broad cultural factors such as Collectivism and Power Distance, but it is possible that these features are also influenced by the political and legal systems of China. Unfortunately, cross-cultural studies of performance appraisal tend to ignore political differences between nations that share common cultural backgrounds (Yu & Murphy, 1993).

Performance Appraisal in Unionized Workforces

There are political, legal, and economic forces involved in the relationship between management and labor, and these can have an impact on the way performance appraisal systems are designed, implemented, and used. Union membership is shrinking, particularly in the private section. The U.S. Bureau of Labor Statistics reported that in 2014, approximately 11% of wage and salary workers were members of unions, and of these a large number were in the public sector (e.g., police, fire, teachers, or administrative workers' unions; approximately 35% of these workers are unionized).[7] Nevertheless, there are still a large number of workplaces in which employees are members of unions, and this membership can substantially influence performance appraisal policies and practices.

In workplaces that are not unionized, employees may have little effective chance of influencing human resource policies, although their input is sought and considered in some organizations. Unions give employees a tool to potentially affect a number of characteristics of their work conditions. We typically think of unions as being primarily concerned with salary and with preserving the jobs of their members, but they can have an important role in shaping other human resource policies, including those that govern performance appraisal.

Although unionized firms are less likely to have a performance appraisal system than comparable nonunionized firms (Ng & Maki, 2008), unions are not in principle opposed to performance appraisal. Rather, they are likely to be opposed to performance appraisal systems that are informal or that appear to be unduly subjective. For example, there is evidence that in unionized settings, performance appraisal processes tend to be more regular and formalized than in comparable settings that are not unionized (Ng & Maki, 2008; Verma, 2005). Unions are more likely to be concerned with the uses of performance appraisal than with appraisal per se. In a union environment, performance appraisal is less likely to be used for salary, promotion, or layoffs (Bennett & Kaufman, 2007).

Finally, unions may take a role (at least a consultative role) in the development of performance appraisal systems. Later in this book, we will review evidence that employees have more positive attitudes toward

performance appraisal if they have some input or voice in the development of these systems and in the evaluations they receive, and a union can be an effective method of giving employees this voice. In addition to giving employees a voice, union involvement may help safeguard organizations from making poorly considered changes to performance appraisal systems to reflect the latest management fad or craze.

Economic Environment

Changes in economic conditions can lead to substantial changes in the way performance appraisals are conducted and used. For example, there is evidence that during economic downturns, organizations tend to favor human resource systems that maximize managerial control and predictability (Maley, 2013). For example, this might mean more emphasis of top-down evaluation and less on participative appraisal systems.

Although research on the growth and decline of organizations is not concerned solely with the economic environment, this literature is nevertheless the source for several hypotheses about the relationship between the state of the economy and the methods used to evaluate workers' performance. Early in their life cycles, organizations are often characterized by liberal goals and flexible structures. However, as organizations prosper and grow, they tend to become inflexible and autocratic (Whetten, 1987). As organizations decline, relationships within the organization become increasingly politicized (Pfeffer & Salancik, 1978). Decline leads to higher levels of conflict, secrecy, and scapegoating, and to lower levels of morale, innovation, and participation (Whetten, 1987). We might expect similar phenomena to occur during either highly prosperous (growth) or economically unfavorable (decline) times.

As organizations grow and prosper, we would expect that performance appraisal will become increasingly formalized. In part, this is a product of the size of the organization; as the organization gets larger, informal relations between workers at different levels become less likely. As organizations decline, you might find more emphasis on documentation of poor performance, especially if the organization is forced to lay off a significant portion of its workforce. We might also expect less emphasis on feedback and development (as a result of the increasing emphasis on secrecy), and more interest in using performance appraisal as an instrument to increase production.

Negative economic trends probably have a greater impact on behavior in organizations than do positive ones. In particular, personnel practices that are adopted during hard times are likely to be retained when the economy rebounds. This is precisely the pattern with union–management relations in the early and mid-1980s. The recession in the early 1980s created conditions where unions were forced to make substantial concessions to management, often in the form of decreases in pay. When the recession ended, few organizations restored their earlier contracts; requests for pay cuts and other major concessions are still common in union–management negotiations.

Similarly, broad economic trends can have effects on labor markets that indirectly influence virtually every HR practice including performance appraisal. For example, during the period 2007–2014, the global financial collapse led to high unemployment in many countries. This, in turn, substantially changed the balance of power between employers and employees, making employees reluctant to change jobs or to raise complaints that might lead them to losing their current jobs. There is evidence of substantial growth in the part-time and contingent workforce, although the role of the financial crisis in this growth is uncertain.[8] What is beyond dispute is that organizations must frequently change their HR systems to adjust to both the different balance of power between management and labor and to the changing nature of the workforce. For example, an increased reliance on temporary and contingent workers is likely to decrease the role of performance appraisal in determining salary raises and promotions, since large sectors of the workforce might not qualify for these. The changing balance of power and influence might encourage more top-down appraisal and less participative systems, on the theory that some workers might have less strong and stable connections to the organization than others.

The Work Environment

Many personnel programs and interventions are built around the assumption that individuals are largely responsible for their high or low performance levels, and that interventions aimed at improving ability and/or motivation should increase performance (Dobbins, Cardy, Facteau, & Miller, 1993). However, poor performance is often the result of situational constraints rather than individual effort or skill (Kane, 1993). Situational constraints include lack of information needed to complete tasks; lack of adequate tools, equipment, material, and supplies; insufficient budgetary support; inability to obtain necessary help from others; lack of adequate preparation; lack of time; and environmental hazards and obstacles (Peters & O'Connor, 1980; Peters, O'Connor, & Eulberg, 1985). Of the 11 categories of situational constraints reviewed by Peters, O'Connor, and colleagues, three represent aspects of the physical environment that might constrain an individual's ability to perform well: (a) shortages of tools and equipment, (b) lack of materials and supplies, and (c) an unfavorable work environment (e.g., poor lighting).

Unfortunately, raters and ratees do not always have consistent perceptions of these constraints, and a ratee who is unable to perform well because of such constraints might receive a negative evaluation from a rater who is unaware of them. On the other hand, situational factors can enhance performance (Cardy, Dobbins, & Carson, 1995). Workers who have an abundance of resources, help from others, limited time pressures, and the like probably perform better than other workers whose ability and effort is similar but whose situations are less bountiful.

Jawahar (2005) suggests that raters do in fact take situational factors into account when evaluating their subordinates, but that some raters are better than others at making appropriate adjustments in their evaluations. In particular, raters who are higher on self-monitoring ability appear to do a better job taking situational constraints into account.

Effects of Situational Constraints on Performance Appraisal

Research has identified a number of consequences of situational constraints. First, there is evidence that situational constraints lead to lower levels of performance (Peters & O'Connor, 1980). This should not be surprising, since situational constraints are defined as variables that constrain one's ability to perform well. Second, it has been hypothesized that the presence of situational constraints will lead to decreased variability in performance, in part because of a ceiling effect. Empirical research, however, has not consistently supported this hypothesis (Peters et al., 1985; Peters, O'Connor, & Rudolph, 1980). As we will note below, the failure to find range restriction effects may say more about the extent to which situational constraints exist than about their effects. Third, situational constraints appear to lead to negative affective reactions and to lower motivation, although research results here are somewhat mixed (Peters et al., 1985; Phillips & Freedman, 1984).

In their review, Peters et al. (1985) note that serious situational constraints are rare. It may be that the small effect sizes found in situational constraint research reflect the fact that work environments are generally munificent, and that the situational constraints that do exist are too minor to have much of an impact. This is another area in which cross-cultural and historical research methods may be worthwhile. We may be underestimating the effects of situational constraints as a result of doing research in a time and place where these constraints are rare. We predict that there would be stronger effects in settings where the seriousness of situational constraints was variable, rather than being uniformly low.

In situations where we cannot remove situational constraints, it will be necessary to develop methods of measuring performance that take these constraints into account. It is hardly fair to compare the performance of someone who has all of the tools, equipment, and supplies necessary to complete the task to someone else who does not. Peters et al. (1985) review the results of several studies that have attempted to develop performance appraisal methods that take situational constraints into account. Unfortunately, the results of these studies have

not been encouraging. To date, no methods have been developed that leads raters to appropriately adjust for the presence or absence of situational constraints.

As you may have noted earlier, some of the effects of the physical environment tend to mirror those of the economic environment. In both cases, we expect that shortages and scarcity will lead to increased levels of conflict, and to increasing politicization of the organization. Scarcity also leads to an emphasis on conservation and efficiency, which may lead to a change in the performance dimensions that are viewed as most important by an organization. The physical environment may, however, have a more immediate and vivid impact. If the tools, resources, and information necessary to do the job are not available, the consequences are felt immediately, whereas poor economic conditions may take a significant period to lead to changes on the shop floor.

Technological Changes and Performance Measurement

The world of work is becoming increasingly complex, and more jobs than ever require both the ability to deal with complex problems and skill in working with technology (National Research Council, 2012; Organization for Economic Cooperation and Development [OECD], 2012). One result is that an increasing number of workers perform in jobs or in settings that make it possible to automatically measure some aspects of their performance through technologies that range from cameras to keystroke monitoring software. We reviewed the topic of electronic performance measurement in more detail in Chapter 5, but it is useful to note here that the mere availability of such methods can change both performance and performance measurement. For example, employees who work using computers belonging to or networked with their organization have good reason to be concerned that their patterns of computer use might be monitored by their employer, and that there could be adverse employment outcomes if their behavior does not match the preferences and regulations of their employer. Even if employers do not engage in monitoring their employees' computer use, the very fact that they can and they might could have an impact on employees. As our discussion Chapter 5 suggested, these effects can be positive or negative, and employers should not assume that electronic performance monitoring is necessarily a good idea.

Proximal Context

There are a number of variables within organizations that can directly influence the way performance appraisal systems are designed, implemented, and used. Levy and Williams (2004) distinguished between proximal variables that influence how performance ratings are carried out (*process proximal variables*) and those that deal with the configuration of the appraisal system itself (*structural proximal variables*). We believe it is useful to expand the definition of process proximal variables to those that influence both how ratings are carried out and how they are interpreted, since many of the proximal context variables that have been studied influence the interpretation of performance appraisal as well as its implementation. Table 7.5 lists examples of process and structural proximal context variables that have been most extensively studied. We will discuss process variables here; structural proximal context variables are discussed in several of the chapters that follow.

Table 7.5 Proximal Context: Process and Structural Variables

Process	Organizational climate and culture
	Organizational strategies
	Work groups
	Supervisor–subordinate relationships
Structural	Who—the targets of appraisal

	When—the timing of appraisal
	By whom—source of ratings/evaluations

Organizational Climate and Culture

Schneider, Ehrhart, and Macey (2013) note that

organizational climate may be defined as the shared perceptions of and the meaning attached to the policies, practices, and procedures employees experience and the behaviors they observe getting rewarded and that are supported and expected… . On the other hand, organizational culture may be defined as the shared basic assumptions, values, and beliefs that characterize a setting and are taught to newcomers as the proper way to think and feel, communicated by the myths and stories people tell about how the organization came to be the way it is as it solved problems associated with external adaptation and internal integration. (p. 361)

DeNisi and Smith (2014) suggest that understanding the climate and culture of an organization can help us understand why performance appraisal succeeds or fails. In particular, shared perceptions (to the extent which they actually are shared) regarding the behaviors that are valued, the policies that are important, and the norms of the organization can help support performance appraisal systems. However, if the shared perception is that performance appraisal is unimportant, or that it is unduly political, these perceptions can undermine even the most carefully constructed system (see, for example, Tziner, 1999). DeNisi and Smith (2014) propose that both national cultures and corporate cultures represent critically important contextual factors that define what types of performance appraisal systems are more or less likely to work.

Defining and Measuring Organizational Culture and Climate

There are many different measures that purport to shed light on the climates and cultures of organizations (Schneider, Ehrhart, & Macey, 2011, 2013), and there is no single dominant model to define precisely what defines the climate or culture of any particular organization. Rather, the literature on organizational climates and cultures has focused on a wide array of quite specific climates (e.g., climates related to safety, service, procedural justice, ethical behavior, or diversity) and on the relationships between specific organizational cultures and dependent variables such as leadership or firm economic performance (Schneider et al., 2013). There have not been many successful taxonomies of organizational cultures; perhaps the most useful way to describe organizational cultures in a general way is through Quinn and Rohrbaugh's (1983) competing values framework.

Quinn and Rohrbaugh (1983) propose that organizational cultures can be characterized in terms of the 2 x 2 framework illustrated in Figure 7.2. This model has received considerable support; its relevance to performance appraisal can be most easily illustrated by considering the types of performance appraisal systems that would be most likely to succeed or fail in organizations that exhibit the four different types of cultures this framework describes.

Organizations whose culture is flexible and internally focused exhibit what this model describes as a Clan Structure, in which the underlying assumption is that people will behave appropriately if they have trust in and loyalty to the organization, and in which the primary values reflect attachment and support. A performance appraisal that attempts to separate employees into different classes (e.g., the "rank-and-yank" system developed at General Electric, in which a set proportion of lowest-ranked employees was designated to be terminated each year and replaced by new hires) will probably fail here, while developmentally oriented appraisal systems (e.g., those that focus on identifying developmental needs) are more likely to succeed.

Organizations whose culture is stable and internally focused exhibit what this model describes as a Hierarchy Structure, in which the underlying assumption is that people will behave appropriately if they have clear, well-defined roles and formally defined policies and procedures to follow, and in which the primary values reflect communication and formalization. Performance appraisal systems that violate these structures (e.g., 360-degree feedback systems in which peers and subordinates give performance feedback that may be as influential as the feedback given by superiors) will probably fail, whereas appraisal systems that respect these structures (e.g., traditional top-down appraisals) may be more likely to succeed.

Figure 7.2 The Competing Values Framework of Organizational Cultures

Figure 7.2 The Competing Values Framework of Organizational Cultures

Organizations whose culture is flexible and externally focused exhibit what this model describes as an Adhocracy Structure, in which the underlying assumption is that people will behave appropriately if they understand the importance and the impact of the tasks they are asked to perform, and in which the primary values reflect growth and autonomy. Performance appraisal systems that are consistent with these values (e.g., developmentally oriented appraisal systems) will be more likely to succeed than appraisal systems that contradict them (e.g., highly formalized systems that restrict evaluation to a predetermined set of dimensions, with little room for employee voice or for individual goal setting).

Organizations whose culture is stable and externally focused exhibit what this model describes as a Market Structure, in which the underlying assumption is that people will behave appropriately if they have clear objectives and are rewarded for their individual accomplishments, and in which the primary values reflect achievement and competence. Appraisal systems that recognize and reward individual differences in accomplishments will be more likely to succeed than systems that treat all employees pretty much the same, such as appraisal systems that are tied to merit pay systems with pay pools so small that there is no room to effectively distinguish truly good from average performers.

There are, of course, alternatives to the competing values framework. In addition to studying national cultures, Hofstede developed an influential model of organizational culture (Hofstede, 2001). He described organizational cultures in terms of: (1) means oriented versus goal oriented—emphasis on how work is carried out versus what outcomes are achieved, (2) internally versus externally driven—extent to which concern for doing what organizations believe is best versus doing what customers prefer, (3) easygoing versus strong work discipline—relative emphasis on control and discipline, (4) local versus professional—extent to which employees identify with the local organization versus their discipline and profession, (5) open versus closed system—accessibility and ease of entry into an organization, (6) employee-oriented versus work-oriented—emphasis on pressure to perform versus the welfare of employees, (7) degree of acceptance of leadership style—extent to which the leadership styles of direct supervisors are consistent with subordinates' preferences, and (8) degree of identification with your organization—extent to which employees feel a direct link with the organization as a whole. Table 7.6 lays out some ways in which organizational differences in these characteristics might influence performance appraisal.

Table 7.6 How the Hofstede Dimension of Organizational Culture Might Influence Performance Appraisal

If the organization is:	Performance appraisal systems might:
Goal oriented	Include detailed measures of whether appropriate procedures were followed
Externally driven	Include input from customers or clients
Easygoing	Be highly informal
Professional in orientation	Focus on development of high-level skills
An open system	Include input from customers or clients
Work-oriented	Focus on objective productivity metrics
High on acceptance of leadership style	Be more likely to be characterized by a higher level of trust
High on degree of identification with your organization	Be more likely to be accepted and trusted

Do All Organizations Have Discernable Climates or Cultures?

The most difficult issue is understanding whether organizations actually have climates or cultures in any meaningful sense. As Schneider et al. (2013) note, the constructs of organizational climate and culture assume some level of shared agreement, and this is not likely to be true in all organizations. Even if an organization has a well-defined climate and culture, the relevance of that climate and culture to performance appraisal might be limited. For example, many organizations whose employees engage in dangerous activities might develop a strong safety culture, which might be largely irrelevant to performance appraisal.

Organizational climates and cultures are most likely to influence appraisal systems when they focus on one or more of the issues described in Table 7.7. For example, if an organization's culture emphasizes equity and transparency and its climate is one in which trust predominates, this will probably have a strong influence on how performance appraisal is conducted and on the success or failure of appraisal systems. Another organization in which competition and building and exercising influence predominates might develop a quite different type of appraisal system.

In Chapter 10, we will review research on reactions to performance appraisal systems in more detail, but it is useful to make the point here that climates that inspire trust (these climates often put an emphasis on transparency and fairness) are much more favorable environments for the development of a successful performance appraisal system than climates that inspire cynicism and political manipulation. In the final two chapters of this book, we will note how successful performance appraisal cannot be achieved in a vacuum, and that there are specific organizational characteristics that make it easier or harder to develop effective performance appraisal systems.

Table 7.7 Aspects of Climate and Culture Most Likely to Influence Appraisal Systems

Equity versus Efficiency—climates and cultures are most likely to influence appraisal systems when they take a strong stance on the question of whether people should be treated the same versus treated in accordance with their individual accomplishments
Trust versus Cynicism—climates and cultures are most likely to influence appraisal systems when they are driven by strong norms of mutual trust or by strong levels of cynicism
Transparency versus Political Intrigue—climates and cultures are most likely to influence appraisal systems when they are driven by strong norms of either making decisions and decision processes as transparent as possible or by reliance on political skill and influence to accomplish desired ends

Organizational Strategy

Businesses can pursue a wide range of strategies in their pursuit of success. Consider, for example, the four different strategies that are shown in Table 7.8 for distinguishing your organization from its competitors. An organization that pursues a growth strategy will place a premium on innovation and development, whereas an organization that follows a cost-containment strategy will place more emphasis on efficiency and reduction of errors. Many businesses follow some mixture of the four strategies listed below, but in general, organizations that pursue different strategies are likely to prefer different methods for evaluating the performance of individuals, work groups, and organizations (Dekker, Groot, & Schoute, 2013). Dekker et al. (2013) review research linking performance measurement methods with business strategy. One of their principal conclusions is that organizations that pursue mixed or hybrid strategies (e.g., using both growth and cost control to maximize their competitive advantage) are likely to need more complex methods of defining and measuring performance than organizations that pursue only one type of strategy. This makes a good deal of sense, because each strategy emphasizes potentially different aspects of performance.

Table 7.8 Four Strategies for Gaining a Competitive Advantage

Growth	Add new products or add new features to existing products
Product Differentiation	Distinguish yourself from competitors through superior products or service
Cost Control	Provide goods and services at a lower price than your competitors
Acquisition	Purchase or take control of potential competitors

Dekker et al. (2013) also note that strategies can have a bearing on the ways performance information is used. As we note in Chapter 8, there are many different uses for performance appraisal in organizations, including administrative uses (e.g., using appraisals in determining pay or promotion) and developmental uses (e.g., giving feedback about strengths and weaknesses). They suggest that mixed or hybrid strategies are often linked to a wider range of uses of performance appraisal than is common among organizations that pursue a single strategy for gaining competitive advantage. They also tested the hypothesis that organizations that pursue mixed or hybrid strategies might employ a wider range of incentive schemes than organizations that pursue a single business strategy, but found only mixed support for this hypothesis.

Work Group and Team Characteristics

Most jobs involve working in groups or teams, at least part of the time, and these groups can influence both the job performance and the supervisor's rating of each group member (Hackman, 1976). First, groups can influence the knowledge and skills of individuals, the level of psychological arousal while working, and the member's effort and performance strategies. A work group can be a source of help and assistance, providing each member with resources he or she might not otherwise have, but it can also be a drag on the performance of group members. Norms or expectations develop within groups regarding acceptable levels of performance, and these norms become stronger as the group increases in cohesiveness. In a classic ethnographic study, Whyte (1943) showed how work group members enforce these norms, shunning or otherwise punishing group members who out-performed the group norm (thus making everyone else look bad).

Not only do norms develop concerning productivity, they also develop to define what are acceptable or unacceptable performance ratings. For some groups, a norm develops that leads raters to believe that high or inflated ratings are the only acceptable ratings. For example, Cleveland, Morrison, and Bjerke (1986) found that when there were high entrance or educational standards for admission into a group or job community, raters were reluctant to evaluate employees as average or below. The rater's rationale was that these individuals passed through numerous performance hurdles in order to gain access to the group, and therefore were all "superior."

The characteristics of a group such as the performance variability of its members (Ilgen & Feldman, 1983; Liden & Mitchell, 1983) and the salience of specific group members (Brewer, 1979; Linville & Jones, 1980;

Quattrone & Jones, 1980) also influence performance ratings. Raters do not evaluate one person on a single occasion; rather, evaluation occurs for many people over an extended time, or covering many tasks. As a result, each group member becomes a point of reference for evaluating other members (Ilgen & Feldman, 1983). For example, Mitchell and Liden (1982) found that in a work group with one poor performer and two good ones, the evaluation of the poor performer was higher than it should have been, and the evaluations of good performers were lower. This was especially true when the poor performer was portrayed as high in popularity and leadership skills.

There is a useful distinction between work groups and work teams. Teams have two defining characteristics, interdependence and shared responsibility for outcomes (Guzzo & Dickinson, 1996; Hollenbeck, Beersma, & Schouten, 2012), whereas a work group might not exhibit interdependence if each member works pretty much on his or her own. In teams, however, the ability of team members to perform their tasks sometime depends on other team members performing *their* tasks well, and the team as a whole is responsible for particular products or outcomes.

Various types of teams exist in organizations, including: (1) project teams/task forces—teams put together to carry out a particular task or project, which are dissolved when the project is done; (2) cross-functional teams—teams whose members have different skills or specialties; (3) virtual teams—teams whose members are scattered geographically, and who work together via electronic communication or some other method of sharing tasks without meeting in the same physical location; and (4) self-managed teams—teams with a flat authority structure, where formal leadership is less important than emergent leadership (i.e., the emergence of some member or members who take on a leadership role even though their formal titles do not give them authority over others). Hollenbeck et al. (2012) suggest that rather than thinking about different *types* of teams, it is more useful to think of teams as varying on three core dimensions: (1) skill differentiation—the extent to which members possess similar or different skills, (2) authority differentiation—the extent to which different team members have formal authority over others, and (3) temporal stability—the extent to which team members have a history of working together and an expectation of continuing to work together in the future.

The existence of teams has clear implications for performance appraisal. First, teams provide an obvious opportunity to use peer ratings in evaluating performance. Team members are likely to be knowledgeable about each other's performance and contribution to the team effort, and they have a vested interest in improving each other's performance. There is evidence, however, that team members are sometimes reluctant to evaluate each other, and when forced to do so, may resist differentiating between members in terms of their performance. This tendency is far from universal, but in teams that decline to differentiate, team members report higher satisfaction and stronger perceptions of procedural justice (Drexler, Beehr, & Stetz, 2001). Whether these attitudes are the cause or the effect of an unwillingness to differentiate among team members when rating their performance is unclear, and arguments can be made for either causal flow.

Second, the existence of teams probably changes the content of appraisals. For example, when switching from evaluations of individuals working alone to evaluations of team members, the relative importance of task performance versus organizational citizenship is likely to change. Harris and Barnes-Farrell (1997) identified several core components of teamwork (e.g., team leadership monitoring, communication, feedback, backup behavior, coordination) and showed that all of these were related to evaluations of the effectiveness of team members. This makes perfect sense, because all of these behaviors are likely to make the team as a whole effective. In teams where the level of interdependence is high, team members who exhibit high levels of behaviors that support the team as a whole might be viewed as good performers even if their individual task performance might not be not so good.

Finally, the three characteristics of teams identified by Hollenbeck et al. (2012) have clear implications for performance appraisal. For example, when a team is high on skill differentiation, one implication is that the content of appraisals (i.e., the performance dimensions that are evaluated) will vary from member to member, making it hard to directly compare the performance of different team members. Consider, for example, a type of

team that has been extensively studied: a flight crew (Guzzo & Dickinson, 1996). The iconic U.S. bomber aircraft, the B-52, has a crew that typically includes a pilot, co-pilot, navigator, radar navigator/bombardier, and electronic warfare officer. Because these team members have distinctly different roles and responsibilities it can be hard to determine whether the co-pilot performs better at his job than the navigator does in his particular role.[9]

A team that is high on authority differentiation will probably require different strategies for performance appraisal than a team where there are few differences in formal authority; peer evaluations are more likely to be acceptable in teams with a flat authority structure. Finally, teams that are long-lived will probably be evaluated in different ways than teams that are put together for a specific purpose and disbanded soon afterward. Teamwork is probably a more critical performance dimension for members of teams that are going to be together for extended periods than for short-lived task forces. In a short-lived team, people can probably get along for a limited period even if they do not possess well-developed teamwork skills.

Supervisor–Subordinate Relationships

The nature of your relationship with your supervisor has a strong influence on the performance ratings you receive. For example, Duarte, Goodson, and Klich (1994) showed that regardless of objective measures of performance, ratees who experience high-quality leader–member relationships receive higher performance ratings. Of course, ratee characteristics such as competence and dependability affect the nature of the leader–member relationship, and someone who is a perennially poor performer is unlikely to attain or remain in a high-quality relationship with his or her supervisor. Nevertheless, employees who have established a pattern of good performance and who maintain a strong relationship with their supervisor may receive high ratings even if their performance during the period the ratings covered is in fact not so good.

Supervisor–subordinate relationships not only influence ratings, they also influence the type of feedback ratees receive. Among the contextual factors thought to influence the likelihood of seeking and receiving feedback, the relationship between supervisors and subordinates has received the most attention (Lam, Peng, Wong, & Lau, 2015). Leader–member exchange (LMX) theory proposes that leaders have qualitatively different relationships with employees who are members of in-groups versus out-groups (Dansereau & Graen, 1975; Graen, 1976; Graen & Uhl-Bien, 1995). In-group employees are those employees who are trusted and valued by the leader and who often have a stronger connection with the leader, whereas out-group employees are those employees whose performance, motivation, or ability are not highly regarded by the leader and who have more formal and perhaps even adversarial relationships with their leaders. The defining feature of the in-group is trust, which makes it easier to give and receive feedback. On the other hand, leaders often value and depend on members of the in-group, and may be reluctant to give them negative feedback, even when it is deserved.

There is considerable evidence that a high-quality relationship between leaders and followers encourages feedback seeking and that a low-quality relationship encourages feedback avoidance. Moss, Sanchez, Brumbaugh, and Borkowski (2009) showed that the likelihood of feedback avoidance can explain the relationship between relationship quality and subordinate performance. That is, it is often the case that employees who have a low-quality relationship with their supervisor avoid and ignore performance feedback, which in turn makes them less effective at their job. This is not to say that feedback avoidance is the only mechanism by which LMX quality influences subordinate performance, but it does indicate that the relationship between LMX quality and feedback avoidance is sufficiently strong to make this sort of mediation possible (see also Chen, Lam, & Zhong, 2007).

Finally, the physical arrangement of the workplace itself can affect the nature of the supervisor–subordinate relationship in a way that influences performance appraisal. The physical distance between raters and ratees and between ratees is important in supervisor–subordinate interactions and evaluations because it can affect the opportunity to observe behavior and performance (Wexley & Klimoski, 1984). The less your supervisor sees of

your performance, the less likely it is that you will be a member of his or her in-group, and this could influence both the performance ratings and the feedback you receive.

The Implications of Context Dependence on Transferring Policies or Practices Across Contexts

The ways performance appraisal systems are designed, implemented, interpreted, and experienced are all influenced by the organizational and social context in which these systems exist (Judge & Ferris, 1993; Mitchell, 1983; Murphy & Cleveland, 1995). Both proximal contexts, such as the climate and culture of the organization and the composition of work teams and departments and distal contexts, such as the state of the economy or the nature of the legal system influence performance appraisals. For example, performance appraisal is probably used and interpreted differently during good economic times (when organizations might have the ability to reward good performance) than during economic downturns (where there are few rewards available). Performance appraisal is conducted and used differently if organizations are worried about the potential for legal scrutiny (e.g., because of a pending discrimination charge or the possibility of being investigated) than when they are not. Performance appraisals are conducted and used differently in organizations with supportive and open climates than in organizations where suspicion and cynicism abounds (Bernardin, Cooke, & Villanova, 2000; Tziner, Murphy, & Cleveland, 2005; Tziner, Murphy, Cleveland, Beaudin, & Marchand, 1998).

The practical impact of this context dependence is that it can be very difficult to predict whether or not a particular performance appraisal system will work when it is transferred to a new context. A system that seems to works well in one culture, organization, or setting might fail absolutely in another. A system that seemed to work well one year might fail terribly the next year. The lack of transferability of performance appraisal systems across organizations or across contexts is a serious problem, especially because of the well-known strategy of organizations to attempt to copy human resource management systems and innovations that seem to be successful in prestigious companies (DiMaggio & Powell, 1983). For example, during the highly praised tenure of Jack Welch as CEO, General Electric introduced a "rank and yank" system of evaluation in which employees were sorted into a small number of categories, and in which employees in the lowest category (typically the lowest 10%) were liable to be dismissed or laid off. Because of the prestige of Welch and General Electric, this type of system spread into a number of other companies, regardless of whether it fit the missions, climates, cultures, or circumstances of those companies. At least in the field of performance appraisal, this tendency to try and copy the policies and practices of prestigious companies is unlikely to work, in part because the new context in which those policies are applied is unlikely to be identical to the context in which these systems or policies were developed.

Summary

Performance appraisal occurs in a specific context that influences virtually all aspects of the appraisal process. First, there is the broad culture in which appraisals are done. These cultures vary on a number of characteristics, including Collectivism–Individualism and Power Distance, which can have a strong influence on the type of appraisal system that is developed, the way it is applied in organizations, and the way appraisal information is used. Next, appraisal occurs in a specific legal and political context. For example, laws designed to encourage equal employment opportunities strongly influence the way appraisals are conducted in the United States, pushing American performance appraisal systems in the direction of careful documentation of the job-relatedness of appraisals and of the fairness and freedom from bias of performance evaluations.

Appraisals are influenced by the economic climate. For example, during tough economic times organizations may need to focus on minimizing costs and errors and on maximizing short-term profitability. In a more relaxed economic environment, this same organization might be able to invest more time and energy in employee

development. Appraisals are also influenced by the physical and technological environment. When workers lack the resources, information, tools and equipment, or materials needed to perform their jobs, performance appraisal might be pointless; unless employees have at least a *chance* of success, there is little to be learned by measuring their individual performance.

Much like national cultures, organizational climates and cultures influence appraisal. The extent to which the climate and culture of an organization encourages trust and involvement versus cynicism and political manipulation is likely to be highly important. Organizational strategies also make a difference. An organization that focuses on cost control will probably develop a quite different appraisal system than one that attempts to produce the highest-quality products.

If we drill down to contextual variables within organizations, two stand out as most critical. First, the nature of the work group or team can have a substantial impact on the way appraisals are conducted and the way they are interpreted. Similarly, the relationship between the supervisor and subordinate will be an important driver of performance ratings and performance feedback.

The most general message of this chapter is that performance appraisal can only be properly understood within its particular context. The same exact performance appraisal system might work quite differently in organizations in different countries, or during different sorts of economic times or in organizations that pursue different strategies. For most of the history of scientific research on performance appraisal, context has been largely ignored, but in the last 20 years there has been a substantial surge of interest in understanding how the contexts in which performance appraisals are carried out influence the appraisal process and outcomes.

Case Study: Why Rank and Yank Failed at Microsoft

Many major corporations have adopted some variation of the GE "rank and yank" system in which employees are ranked or sorted into groups based on their performance, and in which the employees who are in the lowest group are under some real pressure to improve or to face meaningful consequences. Technology companies, notably Microsoft, have tried this system, with generally unsatisfactory results, and many are abandoning the idea (Olson, 2013). Why?

Different explanations might be needed to fully capture the reasons why appraisal systems that rely on ranking or stacking employees do not work well in technology firms, but in the case of Microsoft, Olson (2013) argues that it is culture that is to blame. In particular, she notes that Microsoft relies heavily on an information-sharing culture, in which people are willing to take risks and willing to help others succeed. A rank-and-yank system pits employees against one another, and helping a coworker accomplish his or her tasks can directly can make you look worse. Conversely, if other workers fail, you will look relatively good even if your performance is mediocre. Microsoft determined that cooperation and information sharing was more important than the benefits they might expect from a rank-and-yank system.

Even GE, the originator of this system, is abandoning rank and yank, and once again, culture seems to be the culprit. As Nisen (2015) notes, the norms, values, and expectations of workers are different now than they were when Jack Welch first installed the rank-and-yank system, and millennial employees, who prefer frequent and immediate feedback and who value information more than evaluation, are simply not willing to work under a system that pits employees against one another. As a result, GE has moved to a system with a much stronger developmental orientation.

In both cases, the failure of a system that appeared to be highly successful at GE in the 1980s appears to be the result of a mismatch between the demands of the system and the culture of the organization and the society in which the organization is embedded. At Microsoft, the decision to abandon rank and yank was driven by the

differences in the culture of Microsoft and GE. At GE, the decision to abandon rank and yank was driven by the differences in the GE workforce of the 1980s and 1990s and the current GE workforce.

Notes

1. http://dictionary.cambridge.org/dictionary/english/context (British usage).

2. 360-degree feedback systems are discussed in Chapter 9.

3. http://geert-hofstede.com/indonesia.html

4. Over time, the use of age as a criterion for setting wages and evaluating contributions to the organization in Japanese firms has declined.

5. See also Werner and Bolino, 1997.

6. There are some rules and regulations that are common across nations within the European Union, but national legal and regulatory systems are still more important than EU-wide regulations in determining human resource policies.

7. http://www.bls.gov/news.release/union2.nr0.htm; to illustrate the decline in union membership, this same report noted that in 1983 over 20% of workers were members of unions.

8. http://www.cbsnews.com/news/temp-work-raises-long-term-questions-for-economy/

9. We use the masculine pronoun here because, at present, B-52 crews are exclusively male.