For decades, one of the primary goals of organizational research has been the improvement and management of organizational performance. Inherent to the goal of improving performance is the concept of performance measurement (PM). PM is the mechanism that allows managers and researchers to gain an understanding of individual, team, and overall organizational performance. Without the ability to accurately measure a construct such as performance, it is impossible to truly understand, control, or improve it. As Sink and Tuttle (1989) asserted, one cannot manage what one cannot measure. Ultimately, the effective training and management of employees, teams, and organizations in any context is contingent on the quality of PM. Accordingly, much effort has been devoted over the past several decades to exploring theories, methods, and practices associated with PM (e.g., Bititci, Turner, & Begemann, 2000; Campbell, McCloy, Oppler, & Sager, 1993; Folan & Browne, 2005; Gershoni & Rudy, 1981; Kendall & Salas, 2004; Pun & White, 2005).

The PM literature can generally be categorized into three distinct perspectives: individual-level PM, team-level PM, and organizational-level PM. Very little research has simultaneously examined multiple levels. This is problematic given that actual performance in organizations takes place at all three levels simultaneously, and perhaps more important, all three levels of performance are intertwined. Teams are becoming the predominant method for achieving organizational goals. These teams are made up of individual employees, who actually engage in behaviors that lead to performance. Thus, there is a need to integrate these three streams of PM research into one comprehensive understanding of PM and its implications.

To address this need, this chapter presents a multilevel perspective on the field of PM. First, we discuss the criterion problem, which represents a broad issue underscoring the importance of PM. Next, we briefly describe five critical considerations when choosing or designing any PM system. Then, after the core underlying issues are clear, we dive into PM as described from the individual, team, and organizational perspectives. This includes the general definition of performance, key theories, and common measurement strategies used in each stream of literature. Once each perspective is discussed separately, we discuss a multilevel approach to PM. The chapter concludes with a review of current trends requiring future research and some concluding remarks. (See also Vol. 2, chap. 9, this handbook.)
have noted the necessity of well-developed criteria for measuring individual- and team-level performance as well as evaluation of organizational programs and training initiatives (Schmitt & Klimoski, 1991).

Performance criteria are initially conceptual in nature and are thus defined on the basis of subjective statements of what is considered successful performance. Simply stated, performance criteria represent whatever aspects of performance a certain set of stakeholders have identified as critical. Thus, the selected dimensions of any given criterion measure are based largely on the defined conceptual criteria (Nagle, 1953; Toops, 1944). For example, if a set of stakeholders are conceptually interested in assessing the productivity of a professor, this could be assessed using measures of effectiveness such as number of publications, number of graduate students sponsored, and number of conference presentations. The important aspect of selecting criteria measures is to make sure these measures are rationally linked to the conceptual criteria and are sufficiently covering the criterion space. In other words, do the measures of performance effectiveness include all of the things that stakeholders deem important to performance in a particular job?

One of the most troubling issues in performance research has been the lack of focus on the conscious choice and development of criteria measures. Unfortunately, organizations and researchers often select criteria on the basis of availability or how easy the criteria are to collect. This is problematic because the choice of a performance measure influences how well selected predictors can actually forecast future performance. The choice of outcome criteria is just as important as the choice of predictors, if not more so. The best selection test or training system in the world could be developed; however, without sound criteria to serve as a measure of effectiveness, it is difficult to provide evidence for its validity. If the performance measure chosen as the criterion is not conceptually related to the outcome of interest, or if the measure is contaminated, the selection process may be excellent or the training may be well designed; however, there will never be a demonstrated connection to performance.

Therefore, there is a need for an increased focus on developing sound performance criteria and systematically linking those criteria to other constructs of interest. Additionally, performance criteria should be measured without the influence of halo and other sources of error to capture true performance. These issues fall under the term criterion problem (e.g., Flanagan, 1956; Smith, 1976). Essentially, this refers to problems associated with developing and measuring the multidimensional nature of performance criteria given the constraints of the measurement purpose and situational factors (J. T. Austin & Villanova, 1992). For example, Viswesvaran, Schmidt, and Ones (2005) found that less than 10% of the variance in a set of job performance ratings or rankings could be attributed to valid performance-related information. The rest was attributed to such things as halo error, rater leniency error, and random error, among others, demonstrating that these problems are important to consider when measuring performance.

These errors can be divided into three common categories: distributional errors, illusory halo, and other types of errors. Distributional errors relate to incorrect representations of performance distributions across employees being evaluated (Borman, 1991). These errors can occur in both the rating means (e.g., severity or leniency) and variance (e.g., range restriction and central tendency). If a rater provides ratings that are lower (severity) or higher (leniency) than actually warranted by the performance because of inaccurate norms, then ratings will be erroneously deflated (severity) or inflated (leniency). If a rater fails to sufficiently differentiate between two or more raters on the same dimension, then restriction of range has occurred. This is similar to the error of central tendency; however, with central tendency, ratings tend to be clustered around the midpoint of any given scale (Tsui & Barry, 1986). The second category of errors is illusory halo, which results in correlations between ratings of two different dimensions being higher (or lower) than the correlation between the actual behaviors reflecting those dimensions. Essentially, raters are either overestimating (higher correlations) or underestimating (lower correlations) the relationship between dimensions (Borman, 1991; Fisicaro, 1988). The final category of other errors includes such perceptual errors as the similar-to-me error and...
the first-impression error. Similar-to-me error occurs when the rater projects his or her own personal characteristics onto the employee (Latham, Wexley, & Pursell, 1975). If the rater is heavily influenced by early experiences with the ratee, then first-impression error has occurred (Latham et al., 1975). This can cause biased ratings that are lower or higher than the actual performance warrants, depending on whether first impressions are negative or positive.

There are additional issues associated with the development of high-quality performance criteria. Because criteria essentially focus on the results or outcomes of performance, there are several steps required to directly link criteria measurement to associated predictors (J. T. Austin & Villanova, 1992). This is problematic in that other variables, such as situational factors, may constrain this translation from predictors to behaviors to results (Binning & Barrett, 1989). It is important to clearly define the constructs within this context. Borman (1991), in his seminal chapter on job performance, defined behavior (what people do), performance (individual contributions toward organizational goals), and effectiveness (outcomes such as promotion rate or salary level). Campbell et al. (1993) suggested that performance is the actual behavior and therefore measuring the behavior constitutes measuring performance. Regardless of the adopted definition, it is fairly easy to measure behaviors, as they are generally observable and can easily be recorded. It is also a relatively simple process to measure results using quantity, quality, or customer satisfaction. The difficulty lies in (a) tying specific behaviors to specific results in the context of performance and (b) measuring the cognitive aspects associated with behaviors.

Others have suggested that criteria dimensions are also problematic because they are context sensitive (Bailey, 1983). Given this assertion, measures appropriate for use in one situation would be inappropriate within a different context. Also, as noted previously, the selected dimensions of a criterion construct are based on how the conceptual criteria are defined. An additional issue contributing to the criterion problem is the lack of description often provided as to why certain dimensions were selected and other seemingly important dimensions were ignored (J. T. Austin & Villanova, 1992).

SETTING THE STAGE: BASIC CONSIDERATIONS IN PERFORMANCE MEASUREMENT

This section outlines five critical issues to consider when choosing or designing a PM system: (a) the purpose of the measurement, (b) the content of the measurement, (c) the timing of measurement, (d) the fidelity of the measurement setting, and (e) the technique or tools used for measurement. We refer to these issues simply as the why, what, when, where, and how of PM (see Figure 10.1). These considerations are important to keep in mind when examining existing PM strategies, because each strategy presents advantages and disadvantages regarding these considerations. Given that the primary focus of this chapter is on the methods for measuring performance at the individual, team, and organizational levels, the consideration of how (i.e., how to measure performance) is divided into these three perspectives for discussion and represents a large portion of the content in this chapter.

Why: Purpose of Measurement

The first critical issue to consider when choosing or designing a PM system is the purpose for the measurement, as the purpose will drive the entire measurement process (Salas, Burke, & Fowlkes, 2006). The purpose determines whether multiple criteria measures (e.g., Bartram, 2005) or a single composite criterion measure (e.g., Viswesvaran et al., 2005) is used. There are numerous uses for PM data, ranging from basic research to a variety of applied purposes such as training development and strategic planning. The most common purposes for PM include research, feedback development, training development, performance evaluation, and organization planning. Multiple measures are appropriate if the purpose is to diagnose performance issues, as this allows for a more accurate picture of areas needing improvement and aids in planning for training and employee development. Composite measures, on the other hand, are better for comparing across units who may not do the same type of work. This is the basis of the Productivity Measurement and Enhancement System (ProMES), a PM system that is discussed later in the chapter.
Basic research. Accurate PM is absolutely critical to all research endeavors. As stated by Tannenbaum (2006), “measurement lies at the heart of scientific study” (p. 297). Without the ability to accurately and reliably measure performance and other constructs of interest to researchers, it would be impossible to gain any scientific knowledge. Brannick and Prince (1997) also pointed out that “measurement is central to the evaluation and elaboration of theories” (p. 5). Theories would not be validated, or basic relationships tested, without proper measurement. Measurement is the most basic ingredient in any research, for any purpose.

Feedback development. There is a large base of literature connecting feedback to improved performance both for individuals and groups (e.g., Pritchard, Youngcourt, Philo, McMonagle, & David, 2007). PM plays a critical role in the development of feedback. Specifically, performance must be measured to assess how an individual or team is performing, including what they are doing right, what they are doing wrong, and where improvements in performance can be made. These performance data can then be used to develop focused feedback, centered on identified strengths, weaknesses, and areas for improvement. Therefore, accurate and thorough PM is the first step in any feedback system. By accurately measuring and describing the performance of an individual, feedback can be used as specific instructions for performance improvement.

Training development and evaluation. The use of PM data for feedback development is quite similar to the use of PM data for training development. Another intervention designed to improve performance, training, aims to develop an individual’s or team’s knowledge or skills by providing information and opportunities for practice. It is important that training systems are designed to address specific deficiencies in employee performance, as they can often be costly and time consuming both to design and implement. PM data are a necessary first step in the development of training. These data are used to identify deficiencies and pinpoint knowledge, skills, and abilities in need of improvement. PM also plays a role in the assessment of the training system effectiveness. Specifically, performance must be measured at the conclusion of the training program and linked to relevant outcomes to assess whether the training is imparting the desired knowledge or skills (i.e., whether there was learning) and ultimately contributing to the performance of the employees and organization as a whole (i.e., whether there was training transfer). Training is a cycle of providing instruction, assessing learning and outcomes, and adjusting instruction on the basis of that assessment. PM facilitates this cycle.

In addition to remedial efforts, training can also be used to provide new knowledge, skills, or attitudes. For example, nearly 88% of organizations with revenues exceeding $10 billion have executive development programs aimed at providing executives with...
skills or knowledge that are not necessarily directly related to their current positions (Czarnowsky, 2008). PM can aid in the development of these programs as well. By measuring the performance of top executives, an organization can establish criteria of what a successful executive looks like by focusing on the strengths of each individual that positively contribute to personal and organizational effectiveness. These criteria can then be used to develop training targeting future executive development efforts.

**Performance evaluations.** PM data can also be more simply used for evaluation purposes. Employee evaluations are yearly or quarterly assessments used to determine the comparative success of individuals within an organization. Data from these evaluations, usually in the form of subjective ratings performed by supervisors, can then be used to determine various human resources decisions such as promotions, salary changes, or bonuses. PM data for evaluation purposes often serve as the justification behind these types of decisions that must be made in all organizations. Given that performance evaluation data can often impact individual employees in salient and life-changing ways (e.g., firing, promotion), it is absolutely critical that performance data used for this purpose are accurate and nonbiased. Team-level PM can also be used for evaluative purposes, similarly to individual-level data.

**Organizational planning.** Up to this point, every purpose for PM discussed has focused on the individual or team level. However, PM is also critical and necessary at the overall organizational level as well. All organizational-level decision making and planning relies on accurate measurement of performance at the individual, team, and organizational level. For example, if an organization puts a new policy or program in place, they will undoubtedly need to measure performance at some point after implementation to assess whether that program is working as intended and to decide whether the program should be modified, expanded, or eliminated (Tannenbaum, 2006). Assessing big picture performance also allows for an organization to keep track of their organizational health, which can lead to high-level decisions such as mergers or acquisitions. Without PM at the organizational level, decisions such as these would be made blindly. Finally, organizations are interested in overall measures to inform decisions regarding human resources (e.g., recruitment, training).

**What: Content of Measurement**

Once the purpose of the measurement has been identified, it is necessary to determine what corresponding content should be captured. Depending on the reason behind the PM, and how performance is being defined, there are numerous behavioral aspects that could be measured. For example, if the purpose of the measurement is to develop taskwork training for pilots, then measuring task-related performance would likely be the best choice of content. However, if the purpose is to look at how aircrews function together as a cohesive unit, teamwork-related behaviors should be the focus of measurement. As is described later in the chapter, there are many different types of performance that can be measured that focus on task performance, interpersonal performance, or the outcomes of performance. Which type of performance is measured should be decided on carefully to best match the purpose of the PM system. This consideration relates heavily to the criterion problem and the importance of choosing performance measures that represent the conceptual criteria of interest.

**Criteria: A deeper look.** The conceptual criterion can be described as a verbal statement of the important outcomes related to a particular problem (Borman, 1991). Conceptual criteria are abstract statements of what is important to the stakeholder and represent the starting point that drives the development of performance measures. Essentially, conceptual criteria are the gold standard of what a highly successful employee, team, or organization would look like if performing at the highest level. Consequently, conceptual criteria are very subjective in nature. Subject matter experts (SMEs) can provide insight, but the bottom line is that the conceptual criteria should conceptually relate back to the organizational mission and goals. Measures of effectiveness (outcomes) are developed on the basis of the conceptual criteria. These should be developed rationally to ensure they map onto the conceptual crite-
rnia. Measures of effectiveness can be considered the operational definition of the conceptual criteria, with an evaluative component.

Muchinsky (2009) provided an example of conceptual criteria by using the example of a successful college student. He suggested that one important dimension is intellectual growth, noting that highly successful college students experience more intellectual growth than unsuccessful, or less successful, colleagues. Additionally, he pointed to emotional growth as a second dimension, positing that a college education should allow successful students to clarify values and beliefs, aiding in their development and stability. A final dimension to be considered might be citizenship, whereby successful college students desire to engage in civic activities and positively contribute to their surrounding community. Muchinsky suggested that these three factors are the defining criteria, or conceptual criteria, of what constitutes a successful college student. Yet, these are theoretical; therefore, the challenge is to convert these theoretical ideas of desired behaviors into something that can be quantifiably measured.

Performance measures: General characteristics. Generally, performance measures should provide information regarding products, services, or the tasks that individuals or teams complete to produce those products or services. Performance measures are essentially tools that let decision makers see how well individuals, groups, teams, or organizations are doing. Additionally, they provide insight into whether goals are being met, whether customers are satisfied, whether processes are indeed working as desired, and where improvements are needed.

Performance measures can also be multidimensional. There are numerous examples of this dimensionality. For example, number of accidents or injuries per million hours worked is one indicator of a company’s safety program. However, the cost of injuries provides additional information regarding safety program effectiveness. This type of measure provides more detailed information than just the first example, which is a single dimensional measure. Essentially, whatever is measured must be expressed in measurement units that are meaningful given the entire purpose of measurement.

When: Timing of Measurement
Another consideration focuses on when the construct will be measured. Performance is dynamic and changes over time. Processes that are happening at the beginning of a performance cycle may change or even be replaced with different processes at the end of the performance cycle. Therefore, the point during performance at which a construct is measured and the amount of times it is measured (i.e., once or repeated measures) may have a significant impact on what information is captured. For example, many performance measures can be considered “lagging measures” in that they capture performance outcomes long after the behavior that led to those outcomes occurred. End-of-the-year performance reviews and archival data are two good examples of this type of measurement. These sources are practical and useful measures of past performance, and they can be used to link specific performance behaviors to more distal organizational outcomes such as financial success. However, they may provide an inaccurate, or outdated, understanding of current performance, especially if the performance in question is likely to change quickly or often over time (i.e., is cyclical in nature).

Measuring performance throughout a performance period is advantageous because it provides a real-time understanding of what behaviors are actually occurring that lead to the performance outcome. Several of the measures commonly used in the team literature, such as event-based measurement and communication analysis, take this approach. However, the limitation with midperformance measurement is that usually it is a more intrusive method of measurement. If individuals are aware they are being observed or evaluated, there may be an issue with eliciting maximum versus typical performance, which is discussed in more detail in the following section.

One advantage of a repeated measures design for PM is the ability to determine the magnitude of any gains in performance. For example, assume performance is being measured to assess the effectiveness of a training program or some other intervention. A pre- and postmeasurement approach allows for a comparison of levels prior to training and levels posttraining. It is also a useful method for capturing the dynamic aspect of performance. Specifically, dif-
Different performance processes may become more or less critical at different points during a performance cycle, and by measuring performance repeatedly, these changes in process can be captured. However, there are also limitations with this design. Without a control group, it is difficult to conclude with certainty that any noted improvements in performance were specifically due to the intervention and not some outside influence.

Where: Fidelity of the Measurement Setting

Another issue to consider when measuring performance is how characteristics of the setting will impact the process of PM. This issue pertains mostly to observational or rating-type measures, as knowledge tests and financial data are generally independent of the setting. Observational methods, however, are directly influenced by the realism of the measurement setting. In both laboratory and on-the-job settings, the level of fidelity influences the process of PM. Hays and Singer (1989) defined fidelity as “the similarity between the . . . situation and the operational situation which is simulated” (p. 50). In the case of PM, this refers to how closely the measurement setting replicates the actual performance situation it is intended to represent. Fidelity can be further defined in terms of two dimensions: (a) the physical characteristics of the measurement environment (i.e., the look and feel of the equipment and environment) and (b) the functional characteristics of the measurement environment (i.e., the functional aspects of the task and equipment).

Depending on the purpose and nature of the performance measures, different levels of fidelity will be more or less appropriate. For example, if the measurement is intended to capture day-to-day performance of employees on the job, and the job in question is highly dependent on various changes that occur in a fast-paced dynamic environment, a strictly controlled laboratory setting with a low level of fidelity may result in misleading findings. Imagine trying to measure the performance of a team of emergency medical technicians (EMTs) while they are responding to a severe vehicle collision. In this situation, PM may be more accurate if gathered from the natural job environment or in a high-fidelity laboratory setting (i.e., simulation) designed to closely mimic the complex dynamic environment faced by the EMTs. If the simulated environment does not accurately represent the potentially complex environmental factors that are inherent in situations commonly faced by EMTs (e.g., quickly changing medical status of victims, severe weather, vehicle fires or explosions), the measures may not capture performance that is indicative of day-to-day actions. On the other end of the spectrum, some tasks can be appropriately measured using lower fidelity situations. For example, an assembly line worker most likely could perform his or her task in an artificially contrived task simulation in relatively the same manner as he or she performs in the actual work environment. Overall, the level of fidelity of the setting, nature of the task, and purpose of the measurement must be considered in tandem when choosing the setting in which to conduct PM.

Another important issue to consider when examining the fidelity of a measurement environment is the problem of maximum versus typical performance. Maximum performance can be defined as the highest level of performance possible to achieve under optimal conditions, whereas typical performance is the average performance on a day-to-day basis (Mangos & Arnold, 2008). Sackett, Zedeck, and Fogli (1988) provided an example to illustrate the differences between the two constructs, using grocery store register clerks. Typical performance was operationalized as the average number of items scanned less the number of voids per shift, whereas maximum performance was operationalized as the speed and accuracy of scanning items averaged across several timed observations. They found that the measures were not statistically related, which suggests that typical and maximum performance are distinct constructs. Additionally, research has shown that each construct has different antecedents. For example, intelligence is more highly related to maximum performance and personality is more predictive of typical performance (Dubois, Sackett, Zedeck, & Fogli, 1993). There are cultural implications to maximum and typical performance as well. Dubois et al. (1993) found that Caucasians outperformed African Americans on typical performance; however, the differences were minimal when looking at maximum performance.
Certain environmental cues can elicit maximum performance conditions, which can be problematic if the measurement is intended to capture typical, day-to-day performance. When individuals are acutely aware that they are being observed and evaluated, they will likely try to perform to the best of their ability (Sacket et al., 1988). Consequently, maximum performance may be unintentionally elicited and measured. This same phenomenon can occur when observers are present in the natural work environment and the individuals being observed are aware of their presence. This is a significant problem, as research has shown that both maximum and typical performance are predicted by different variables (Campbell et al., 1993; Lim & Ployhart, 2004). Therefore, if the goal of measurement is to represent typical performance, the knowledge of being observed may trigger maximum performance instead and may distort findings regarding the relationships between performance criteria and predictor variables.

Fidelity is often associated with a trade-off in terms of the level of experimental control in a measurement setting. For example, although measurement in the operational work environment is as realistic (i.e., high fidelity levels) as possible, this usually makes it more difficult to isolate and identify the causes underlying performance because so many uncontrolled variables are freely influencing performance. In other words, because the experimenter does not design and control the setting of the measurement, there is the potential for any number of environmental factors to influence or, more important, confound, results. Therefore, it is often more difficult to assess the effectiveness of training or other performance interventions in the field because effects can be hidden by various outside influences. However, results found while measuring performance in the operational setting will most often be more externally valid than results found in more artificial or lower fidelity settings (i.e., lab setting), because data are collected directly in the environment to which they are intended to generalize.

How: Measurement Techniques
One final consideration when choosing or designing a PM system is the technique or approach for measurement. Rather than first presenting the theoretical background for each level, and then the measurement techniques separately, we group them together within the realm of individual, team, and organizational perspectives to ensure that each measurement technique is considered within the appropriate theoretical context. Our hope is that this delineation will provide insight into the most commonly used measurement techniques at the different levels in addition to providing the necessary context for why consideration of multilevels with regard to measurement warrants attention. Therefore, in the next sections, we discuss each level (individual, team, and organizational) and describe common techniques frequently used for PM at that level.

It is absolutely critical to note that the measurement strategies discussed in each perspective are in no way used exclusively within that perspective. Many of the measurement tools mentioned are clearly applicable to, and consequently have been used across, multiple levels of PM. Additionally, the list of measurement strategies we provide is by no means exhaustive. However, as our primary goal is to compare and integrate three distinct streams of literature, we discuss each selected measurement strategy within the theoretical perspective in which it is discussed or most frequently used. We bring all three perspectives together at the conclusion of the chapter.

PERFORMANCE MEASUREMENT FROM THE INDIVIDUAL PERSPECTIVE

In the following section, we discuss PM approaches focused on capturing individual-level phenomenon. First, we define individual performance. Second, we describe several key theories of individual performance that have been developed over the years. Finally, we summarize and describe the most common performance measurement approaches used at the individual level.

Defining Individual Performance
A majority of the PM literature has been devoted to measurement at the individual level. The most basic resource in an organization is the individual employee, and therefore the first place to start man-
Aging performance is at the individual level. For as long as organizational science has existed, the term performance has been misused and vaguely defined. Campbell et al. (1993) attempted to rectify this misuse of the term by defining performance as synonymous with behavior. In this view, performance must be the actions of the individual in question. Job performance, specifically, includes only those actions or behaviors that are relevant to the organization’s goals and that can be scaled in terms of each individual’s proficiency (Campbell et al., 1993, p. 40). Job performance at the individual level, simply stated, is what employees are hired to do.

Therefore, performance is not the outcome or the consequence of behavior; it is the behavior itself (Campbell et al., 1993). This distinction defines the difference between performance, effectiveness, and productivity. Performance is the actions taken by the individual, effectiveness is the “evaluation of the results of performance” and productivity is “the ratio of effectiveness to the cost of achieving that level of effectiveness” (Campbell et al., 1993, p. 41). Each of these terms is an independent construct. One can measure performance without evaluating that performance or without comparing that evaluation with cost. However, when trying to use PM data for any practical purpose, evaluating that performance is a critical step.

Theories of Individual Performance
Several theories of individual performance have been developed looking at various aspects of performance such as job performance, organizational citizenship behavior, contextual performance, adaptive performance, integrated work role performance, and counterproductive work behavior. Each of these theories is described in more detail in this section.

Job performance behaviors. Along with their definition of performance as synonymous with behavior, Campbell et al. (1993) also broke job performance down into eight major behavioral components: (a) job-specific task proficiency, (b) non-job-specific task proficiency, (c) written and oral communication task proficiency, (d) demonstrating effort, (e) maintaining personal discipline, (f) facilitating peer and team performance, (g) supervision or leadership, and (h) management or administration. Job-specific task proficiency reflects the “degree to which the individual can perform the core substantive or technical tasks that are central to the job” (p. 46). Non-job-specific task proficiency is the degree to which the individual can perform tasks in the workplace that are not specific to a particular job (i.e., teamwork skills). Written and oral communication task proficiency is the proficiency with which a job incumbent can write or speak. Demonstrating effort is a reflection of the consistency, frequency, and willingness of an individual to demonstrate effort. Maintaining personal discipline involves the extent to which negative behaviors (i.e., alcohol and substance abuse, excessive absenteeism) are avoided at work. Facilitating peer and team performance is the extent to which an individual supports their peers. Supervision or leadership is the degree to which an individual engages in behaviors directed at influencing the performance of subordinates. Last, management or administration includes performance behaviors directed at management tasks such as articulating goals or monitoring progress. One critical contribution of the Campbell et al. theory of performance is the conceptualization of performance as a multidimensional construct. By breaking job performance down into multiple components, they acknowledged that performance is not just one behavior that can be captured by one simple measure.

Organizational citizenship behavior. One limitation of the Campbell et al. (1993) model of job performance is that it focuses solely on task performance as defined by the job description. It does not account for behaviors that are not technically part of the job yet contribute to job performance. In response to this gap in the literature, several new performance concepts were developed, such as citizenship behavior (e.g., Borman et al., 2001). In a review of organizational citizenship behavior (OCB), Podsakoff, MacKenzie, Paine, and Bachrach (2000) summarized the literature into seven core types of citizenship behaviors: (a) helping behaviors, (b) sportsmanship, (c) organizational loyalty, (d) organizational compliance, (e) individual initiative, (f) civic virtue, and (g) self-development. (See also Vol. 2, chap. 10, this handbook.)

Helping behaviors include helping others with work-related problems as well as actively preventing
problems for others. *Sportsmanship* includes behaviors such as not complaining even when inconvenienced and generally maintaining a positive attitude in the face of difficulty. *Organizational loyalty* is composed of behaviors such as protecting, endorsing, and defending the organization and its objectives. *Organizational compliance* is another organizationally focused type of citizenship behavior that includes internalized and accepting the organization’s rules, regulations, and procedures even when not being directly observed. *Individual initiative* OCBs are behaviors in which the individual goes above and beyond minimum requirements in task-related situations. *Civic virtue* describes a high level commitment to the organization as a whole. This commitment is displayed through behaviors such as attending voluntary meetings and monitoring environmental changes that could impact the organization. Finally, *self-development* refers to voluntary behaviors aimed at improving one’s own knowledge, skills, and abilities.

Overall, OCBs pose an interesting dilemma for PM in that by definition they are not explicit requirements of the job, and thus including them as part of a formal review or performance evaluation may be unethical. Simply stated, it may be inappropriate to evaluate an employee in terms of behaviors that are not explicitly stated as part of their job role, especially if the evaluation is then used as the basis for pay and promotion decisions. If OCBs are included in formal performance reviews, this in essence makes them part of the job description, and therefore they are no longer extrarole. If an organization chooses to include OCBs as part of their official job descriptions, then this approach is appropriate. However, the defining feature of OCBs is that they are performed without being required (i.e., extrarole behaviors; Organ, 1997); therefore, formally measuring them for evaluative purposes could potentially change the nature of these behaviors. In fact, there is an ongoing debate in the literature regarding whether measuring and evaluating OCBs changes the fundamental nature of the behaviors.

**Contextual performance.** The concept of contextual performance is very similar to citizenship behavior in that they both describe on-the-job behavior that is not directly recognized as part of the job (i.e., it is not a job requirement) yet still contributes to job effectiveness. *Contextual performance* is behavior that contributes to organizational effectiveness through its impact on the psychological, social, and organizational context (Motowidlo, 2003). Borman et al. (2001) presented a refined model of contextual performance that categorizes behaviors as *personal support, organizational support, and conscientious initiative.* *Personal support* includes behaviors such as helping others with tasks and showing courtesy and tact when interacting with others. *Organizational support* includes actions such as defending and promoting the organization. *Conscientious initiative* focuses on behaviors such as devoting extra effort to the job or taking advantage of opportunities for self-development. There is a noticeable amount of overlap between the conceptualizations of contextual behavior and OCB as described previously. The same issues regarding measurement of OCBs applies to measuring contextual behavior. Given that contextual performance is composed of behaviors that are not formally recognized as part of the job, it is unethical to evaluate individuals on the basis of contextual performance without their explicit knowledge. Accordingly, if contextual performance is required as part of a job, it is by definition no longer contextual.

**Adaptive performance.** The Campbell et al. (1993) model of job performance also does not account for work behaviors that contribute to effectiveness in dynamic, complex, uncertain, and interdependent settings (Griffin, Neal, & Parker, 2007). Pulakos, Arad, Donovan, and Plamondon (2000) developed a theoretically and empirically based model of performance focused on the concept of adaptivity, with eight dimensions of adaptive performance. This model is intended to assess how well individuals adjust or adapt to new conditions or unexpected job requirements. The eight dimensions of adaptive performance are (a) handling emergencies or crisis situations; (b) handling work stress; (c) solving problems creatively; (d) dealing with uncertain and unpredictable work situations; (e) learning work tasks, technologies, and procedures; (f) demonstrating interpersonal adaptability; (g) demonstrating cultural adaptability; and (h) demonstrating physically oriented adaptability.
The first dimension, **handling emergencies or crisis situations**, involves reacting appropriately in life-threatening or dangerous situations. The dimension of **handling work stress** includes remaining calm when faced with difficulties and effectively managing frustration. **Solving problems creatively** refers to behaviors such as finding innovative ideas to complex problems and considering a wide range of possibilities when solving a problem. **Dealing with uncertain and unpredictable work situations** is similar to handling emergency and crisis situations and handling work stress in that it involved reacting appropriately to a cue, but this dimension focuses on changing plans, goals, and strategies in response to unexpected events or situations rather than just remaining calm. **Learning work tasks, technologies, and procedures** is the most task-relevant dimension of adaptive performance and includes keeping up to date with changing procedures and technology necessary for the job. The dimension of **demonstrating interpersonal adaptability** includes being open-minded and considerate when dealing with other people and maintaining effective relationships. **Demonstrating cultural adaptability** specifically focuses on interacting with people from other cultures and adjusting behavior to make these interactions effective. Finally, **demonstrating physically oriented adaptability** refers to adjusting to physical environmental conditions such as temperature or training to become more physically proficient.

**Integrated work role performance.** Bringing together several of the previous understandings of work performance, Griffin et al. (2007) recently developed an integrated model of work role performance (see Table 10.1). They proposed that context plays a major role in the behaviors that will be viewed as valuable performance in an organization. Specifically, they proposed that uncertainty in the environment influences to what extent roles can be formalized and that interdependence with the environment influences how embedded work roles are in the larger system. Their model attempts to address the difficulty of capturing the total set of performance dimensions in a job by cross-classifying the three levels at which work behaviors can contribute to effectiveness (individual, team, and organization) with the three different forms of work behavior (proficiency, adaptivity, and proactivity). This cross-classification resulted in nine subdimensions of work role performance: (a) individual task proficiency, (b) individual task adaptivity, (c) individual task proactivity, (d) team member proficiency, (e) team member adaptivity, (f) team member proactivity.

<table>
<thead>
<tr>
<th>Individual work role behaviors</th>
<th>Proficiency: Fulfills the prescribed or predictable requirements of the role</th>
<th>Adaptivity: Copes with, responds to, and supports change</th>
<th>Proactivity: Initiates change, is self-starting and future directed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual task behaviors: Behavior contributes to individual effectiveness</td>
<td>Individual task proficiency</td>
<td>Individual task adaptivity</td>
<td>Individual task proactivity</td>
</tr>
<tr>
<td>Team member behaviors: Behavior contributes to team effectiveness rather than individual effectiveness</td>
<td>Team member proficiency</td>
<td>Team member adaptivity</td>
<td>Team member proactivity</td>
</tr>
<tr>
<td>Organization member behaviors: Behavior contributes to organization effectiveness rather than individual or team effectiveness</td>
<td>Organization member proficiency</td>
<td>Organization member adaptivity</td>
<td>Organization member proactivity</td>
</tr>
</tbody>
</table>

(g) organization member proficiency, (h) organization member adaptivity, and (i) organization member proactivity.

**Individual task proficiency** represents the formal task performance that contributes to individual effectiveness. **Team member proficiency** moves one level beyond individual task proficiency and includes task-related behaviors that an individual engages in that contribute to team effectiveness. This includes behaviors such as helping other team members or monitoring the work of other team members. **Organization member proficiency** describes task-related behaviors that contribute to organizational effectiveness, such as defending the organization’s reputation. This pattern of behavior remains consistent throughout the rest of the model. Specifically, **individual task adaptivity**, **team member adaptivity**, and **organization member adaptivity** all refer to behaviors such as appropriately responding to changes in the environment that contribute to individual, team, and organizational effectiveness. Similarly, **individual task proactivity**, **team member proactivity**, and **organization member proactivity** represent the three levels of self-starting, future-directed behavior. This model of performance is more robust than previously mentioned models in that it integrates the broad concepts of role behavior, adaptive behavior, and proactive behavior and applies these concepts across multiple levels of analysis; however, it is still lacking in comprehensiveness (e.g., counterproductive work behaviors [CWBs]).

**Counterproductive work behavior.** Thus far, all theories of individual performance have focused on the positive behaviors job incumbents can engage in. However, humans are also capable of negative, or dysfunctional, work behaviors, and research has labeled this **CWB** (Sackett, 2002). **CWB** refers to any type of intentional employee behavior that is contrary to the organization’s interests. **CWBs** include various deviant acts such as theft, destruction of property, drug abuse, and poor attendance. Some have argued that **CWB** is not a distinct construct but is rather a representation of the negative end of the citizenship behavior continuum. Recently, however, Sackett, Berry, Wiemann, and Laczo (2006) empirically supported that **CWB** is a separate and distinct construct from **OCB**.

**Strategies for Measuring Individual Performance**

The most common approaches for measuring individual performance include performance appraisals, multiple-source ratings, objective measures, job knowledge tests, and work sample tests. It is important to note that these measurement strategies have not been used exclusively for capturing individual performance but are most often seen in the individual realm. Further detail is provided for each strategy in this section.

**Performance appraisals.** Performance appraisals (PAs) are one of the most commonly used methods of PM in organizations. Traditionally, the term performance appraisal referred to a process involving a supervisor completing an annual report on an employee’s performance and discussing it with the employee in an interview (Fletcher, 2001). In a traditional PA system, the supervisor prepares a written evaluation of the employee on the basis of information gathered from coworkers, customers, and any pertinent documentation regarding the employee (Aldakhilallah & Parente, 2002). Then the supervisor schedules a meeting with the employee to review his or her job description, his or her performance against this description, and the organization’s goals. The supervisor also addresses the employee’s career progress and identifies opportunities for further development and improvement. This information is forwarded to higher management to be used in promotion and salary decisions. PA systems are often used to make promotional decisions based on past performance, to identify skill deficiencies and need for training, to make salary decisions, and to provide employees with feedback.

As with most PM approaches, PAs have received both praise and criticism. Some voice the merits of PA systems, including feedback, goal setting, career management, objective assessment, and legal protection (Nickols, 2007). Also, some have claimed quite simply, “having a traditional PA system is better than having no system at all” (Aldakhilallah & Parente, 2002, p. 44), although this claim should be qualified to include only accurate and effective PA systems. It is likely that a bad PA system could actually be worse than no system in that it might force employees to
focus on the “wrong” behaviors—ones that are not really important to effective performance. Another advantage that has been mentioned previously relates to salary—that it is a motivational tool to improve employee performance, and therefore, salary decisions are often tied to the PA system (Rynes, Gerhart, & Parks, 2005). Finally, the outcomes of PAs provide the necessary justifications for decisions such as termination, promotion, transfer, or a change in salary.

The more supported vein of thought, however, posits that supervisor-based PA systems are costly in terms of both time and money, and usually only result in negative emotions with very little demonstrable value (Nickols, 2007). Because the system is almost entirely controlled by the supervisor, there is a high chance that the employee’s appraisal will be based on the supervisors’ opinion alone. Depending on the ethical nature and personality of the supervisor in question, this could make PAs unfair or inaccurate. Data from an Internet survey revealed several other issues associated with PAs (Nickols, 2007). Often, PAs are accompanied by periods of reduced productivity; heightened negative emotional states such as anxiety, depression, and stress; and lowered morale and motivation. When they are linked to short-term rewards or consequences, they also tend to foster a focus on short-term goals at the sacrifice of long-term goals, which can result in negative consequences in the long run. Additionally, people are often hesitant to convey negative feedback, and therefore supervisors may engage in avoidance, delay, or distortion of PA feedback. Benedict and Levine (1988) found that, in particular, female raters may delay appraisals, delay scheduling feedback sessions, and more positively distort their ratings, especially when rating low performers. Because of these inherent issues, much of the research on PA has focused on making more objective and accurate ratings (Fletcher, 2001).

Multiple-source ratings. Multiple-source ratings, also known as “360-degree feedback,” have been defined as “evaluations gathered about a target participant from two or more rating sources, including self, supervisor, peers, direct reports, internal customers, external customers, vendors, or suppliers” (Dalessio, 1998, p. 278). This PM method extends the PA concept by retaining subjective evaluations as the main form of measurement, but this time including ratings from multiple sources, rather than from the supervisor alone. 360-degree feedback systems were originally developed for purely developmental purposes, with no intention of evaluative use. Very quickly, however, organizations began integrating this method into their PA systems, making it evaluative rather than only developmental.

The use of 360-degree feedback at first seems like a logical choice for evaluative purposes (Waldman, Atwater, & Antonioni, 1998). This measurement system, unlike a traditional PA system, provides performance feedback from not only the supervisor but also from subordinates, peers and coworkers, clients (if applicable), and the self, reducing the chances for bias or unfair evaluations (Waldman et al., 1998). Specifically, by having peers rate the performance of other workers at their level, they are in a position to have a deep understanding of the job requirements and conditions, and therefore should provide more accurate ratings. They also generally have more opportunities to observe and monitor the work of the ratee because they work directly with them. Supervisors often are too far removed to have this level of understanding. On this basis, peer-report ratings seem to be a good alternative to self-report ratings because the tendency for inflating performance is reduced.

Additionally, the traditional PA system flows only downward from supervisors to subordinates; supervisors and higher management do not receive any feedback or evaluation. Having feedback flowing in all directions allows for the development of management and leadership, and subordinates and customers are in a good position to evaluate managerial performance (Morgeson, Mumford, & Campion, 2005). 360-degree feedback is also intended to be given anonymously, which theoretically should result in more honest (and therefore accurate) feedback (Ghorpade, 2000). A review of the literature by Morgeson et al. (2005) delineated several other advantages of 360-degree feedback, such as an increase in information and formal feedback between employees, an increase in management learning, encouragement of goal setting and skill development, a change in corporate culture, and improved managerial effectiveness.
However, 360-degree feedback has limitations as a performance evaluation approach. Waldman et al. (1998) suggested that the bidirectional nature of the system could encourage employees to deliberately try to sabotage the system, with supervisors striking deals with their subordinates to give high ratings in exchange for high ratings. They also contended that this type of sabotage is much less likely if feedback is purely developmental, as there is no immediate or direct tangible outcome associated with good or bad evaluations. Also, peer-report ratings come with problems of their own. Specifically, Viswevaran et al. (2005) found that peer ratings had more halo error than supervisor ratings.

Other issues arise regarding the improper implementation of 360-degree feedback systems. Often, organizations implementing the process do not clearly define the mission and scope of the program beforehand or do not provide clear rules for information sharing. Consequently, feedback can end up unrelated to actual employee performance, and employees may be unable to use the results to develop goals or plans (Ghorpade, 2000). Toegel and Conger (2003) argued for the separation of developmental and evaluative tools because of the contradictions and competing goals between using 360-degree feedback as developmental versus an evaluative technique.

Objective measures. Another type of measure often used to assess individual performance is objective data. Specifically, indices such as absences, production rates, sales, or number of disciplinary cases can be used as a measure of an individual’s performance (Borman, 1991). Objective measures are appealing to many organizations because they are easy to gather and interpret, and are not as vulnerable to rater error or subjectivity as subjective measures. However, there are disadvantages to objective measures of individual effectiveness as well. To begin, the quality of an objective measure depends on what the stakeholder in question considers important. If the measurement system is designed to reduce absenteeism, then measuring absences is an appropriate choice. However, if the measurement system is intended to improve customer satisfaction, individual absences would be a very deficient measure. Another important factor to consider with regard to objective measures is the notion of controllability. Controllability of measures can be conceptualized as the extent to which individuals (or teams) control the indicator of performance by varying the amount of effort they allocate to the measured tasks. Controllability—both the actual level and the perception of control—can impact motivation, performance, and ultimately organizational effectiveness. For example, if a measure of performance is bed utilization in a hospital, doctors and nurses have very little control over this, as it is largely determined by the nature of the illnesses that affect the patients. The only way to affect this measure is for patients to remain longer than needed to “use a bed.” This is clearly not an effective practice. Therefore, it is critical that objective measures of individual performance are used only when they appropriately represent the criteria of interest to the stakeholder and they are under the control of the individual being measured.

Job knowledge tests. Job knowledge tests are usually written or computer-based tests that assess the extent of an individual’s knowledge regarding the content and procedures necessary for the job (Borman, 1991). Prior to developing a job knowledge test, a thorough job analysis should be conducted to gain a clear understanding of the knowledge, skills, and abilities necessary to perform the job. From this information, a set of items can be developed. As in any other written test, items can take many forms, such as multiple choice, true–false, or essay. Job knowledge tests are inherently best suited for positions that require high levels of declarative and procedural knowledge. It is important to note that job knowledge tests assess only the extent to which an individual can recall the appropriate information or procedure for a job but not his or her skill in applying that knowledge or performing that procedure. Job knowledge tests can therefore be appropriate performance measures when combined with other indicators (i.e., work-sample tests; see below) or for jobs that are highly dependent on declarative knowledge (e.g., tour guides).

Work-sample tests. Work-sample tests are the practical, organizationally based equivalent of a laboratory-based measurement system (Cascio & Phillips, 1979). Because hands-on work-sample tests
require employees to engage in a simulated version of their normal taskwork specifically for measurement purposes, they are best suited for jobs with tasks that can be easily replicated in artificial settings. When broadly defined to include assessment centers, this type of measurement is suited for many different jobs. Example jobs that commonly use work-sample tests include baggage screeners, assembly line workers, aircraft pilots, or business executives. These jobs include tasks that can be easily simulated or tasks with behaviors that can be easily recorded and scored. Jobs that would not be as well-suited for work-sample tests are more complex or ambiguously defined strategic planning or research positions, as these jobs require tasks that occur over a much longer period of time and tend to include aspects of performance that cannot be outwardly observed as scored during a short, simulated session.

Summary
Several of the most commonly used measures of individual performance were described, including PAs, multiple-source ratings, objective measures, job knowledge tests, and work-sample tests (see Table 10.2 for a summary). These measures capture a wide range of performance behaviors and outcomes from a variety of perspectives. However, to reiterate, the measures discussed do not represent an exhaustive list of the techniques for assessing individual performance. They represent only a sampling of the most commonly used and studies techniques.

PERFORMANCE MEASUREMENT FROM THE TEAM PERSPECTIVE
The science of teams has developed rapidly in recent years, and developments in the measurement of team performance have increased as well. Before we describe the measurement strategies most commonly used to capture team performance, we define team performance and provide a summary of the key theories regarding team performance. (See also chap. 19, this volume.)

Defining Team Performance
“Teams do not behave, individuals do” (Zalesny, Salas, & Prince, 1995, p. 99). This assertion creates a challenge for teams researchers in that it assumes teams do not engage in measurable behaviors. However, it can be argued that although teams do not behave, the behavioral interactions between team members and the behavioral processes the team engages in as a whole can be measured at the team level or aggregated to the team level from the individ-

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance appraisals</td>
<td>Traditionally, a measurement process involving a supervisor completing a written evaluation of an employee’s performance</td>
<td>Aldakhilallah and Parente (2002); Arvey and Murphy (1998); Nickols (2007)</td>
</tr>
<tr>
<td>Multiple-source ratings</td>
<td>Evaluations about a target participant collected from two or more source ratings such as self, supervisors, peers, customers; also referred to as “360-degree feedback”</td>
<td>Beehr et al. (2001); Ghorpade (2000); Morgeson et al. (2005); Toegel and Conger (2003)</td>
</tr>
<tr>
<td>Objective measures</td>
<td>Objective indices such as absences, production rates, sales, or number of disciplinary cases</td>
<td>Borman (1991); Ghalayini and Noble (1996); Pandey (2005); Paranjape et al. (2006)</td>
</tr>
<tr>
<td>Job knowledge tests</td>
<td>Written or computer-based tests that assess the extent of an individual’s knowledge regarding the content and procedures necessary for the job</td>
<td>Osborn and Campbell (1976)</td>
</tr>
<tr>
<td>Work-sample tests</td>
<td>Hands-on simulated versions of everyday taskwork used for measurement purposes</td>
<td>Cascio and Phillips (1979)</td>
</tr>
</tbody>
</table>
ual level. Various models of team performance have identified critical behaviors that facilitate team performance, such as communication, coordination, mutual performance monitoring, and backup behavior (e.g., Marks, Mathieu, & Zaccaro, 2001; Salas, Sims, & Burke, 2005). Volumes have been devoted to the development of behavioral, team-level measurement systems (e.g., Brannick, Salas, & Prince, 1997).

When measuring team performance, it is critical to first start with a definition. Salas, Stagl, Burke, and Goodwin (2007) noted that team performance is often considered either a behavioral or cognitive act, and it is the resulting outputs that are considered performance outcomes. They suggested that team performance is multilevel, characterized by both taskwork (e.g., writing a paper) and teamwork (e.g., backup behavior) competencies exhibited by one or more team members, as well as team-level action (e.g., adaptation). This is not a new conceptualization. Early research has pointed to the multilevel nature of team performance as well (e.g., Kozlowski & Klein, 2000). Salas et al. (2007) postulated that team performance is a bottom-up emergent process, beginning with individuals and progressing toward teams. However, it is important not to ignore the top-down process inherent in team performance. Higher level (i.e., organizational) factors can directly impact, or have a moderating effect on, team performance.

Additionally, it is important to acknowledge that process measures are distinct from, although related to, outcome measures. Process measures capture and “describe the strategies, steps, or procedures used to accomplish a task” (Smith-Jentsch, Cannon-Bowers, Tannenbaum, & Salas, 1998, p. 62). Outcome measures “evaluate the quantity or quality of the end result of those processes” (Smith-Jentsch et al., 1998, p. 62). Although outcome measures are ultimately what the researcher or organization is interested in predicting or managing, it is necessary to measure both process and outcome. This is due to the fact that outcome measures are influenced by much more than just individual or team performance. Outcomes can be affected by environmental factors, situational factors, or even luck.

Cannon-Bowers and Salas (1997) suggested that team PM should capture both processes and outcomes at the individual and team levels. They posited that outcome measures are not very diagnostic because they often do not indicate the underlying causes of that outcome. This is where process measures provide additional information by capturing the underlying behavioral mechanisms of performance. Therefore, it is critical to measure processes as well as outcomes for a robust and comprehensive understanding of team performance.

Theories of Team Performance
Salas et al. (2007) conducted a thorough review of the literature to determine what is known regarding team performance. They found over 130 models of team performance or effectiveness that addressed at least three constructs believed to be relevant in the nomological network of team performance. The inclusion criteria prohibited inclusion of the thousands of additional “models” that only considered one or two constructs. They reviewed 11 of these models (e.g., Gersick, 1988; Hackman, 1987; Nieva, Fleishman, & Reick, 1978), noting that their selection in no way invalidated the significance of the remaining models. On the basis of this review, Salas et al. (2007) created an integrated model of team effectiveness. This model attempted to provide a comprehensive snapshot of the variables that impact team effectiveness.

Below, we use a similar approach in summarizing the team performance literature. Space precludes us from thoroughly examining all potential models of team performance; therefore, we highlight just a selection of the plethora of team performance models. However, through the above discussion, we wish to illustrate that there is no universally accepted model of team performance to use as a departure point for measurement purposes.

Team processes. Capturing performance at multiple levels is also a critical part of the framework presented by Cannon-Bowers and Salas (1997). Specifically, they believed measures of both individual- and team-level competencies are necessary for team PM. Teams are made up of individuals, and teamwork is composed of individual behaviors, so individual-level measurement is critical. However, individual-level measurement is not sufficient, as team performance is more than the
sum of its parts. There must also be team-level measurement to capture the processes and outcomes emerging from team interactions.

Marks et al. (2001) recently delineated 10 core team processes that occur throughout the performance cycle of a team, such as mission analysis, coordination, and conflict management (see Table 10.3). Completing the circle, the outcomes of team performance are also multidimensional. Prior research has examined performance outcomes such as SME ratings of operational readiness in experienced military battalions (Lim & Klein, 2006), performance of undergraduate students in a low-fidelity flight simulation (Mathieu, Hefner, Goodwin, Salas, & Cannon-Bowers, 2000), and safety and efficiency in air-traffic controllers (Smith-Jentsch, Mathieu, & Kraiger, 2005). The definition, and measurement, of team performance varies widely across different populations, settings, and purposes. With so many interrelated inputs, processes, and outcomes, it is impossible for one measurement tool to capture every aspect of team performance simultaneously.

Team performance is also a dynamic phenomenon, which means that static measurement systems

<table>
<thead>
<tr>
<th>Process</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transition processes</strong></td>
<td></td>
</tr>
<tr>
<td>Mission analysis formulation and planning</td>
<td>Interpreting and evaluating the team’s mission, including identifying its main tasks as well as the operative environmental conditions and team resources available for mission execution</td>
</tr>
<tr>
<td>Goal specification</td>
<td>Identifying and prioritizing goals and subgoals for mission accomplishment</td>
</tr>
<tr>
<td>Strategy formulation</td>
<td>Developing alternative courses of action for mission accomplishment</td>
</tr>
<tr>
<td><strong>Action processes</strong></td>
<td></td>
</tr>
<tr>
<td>Monitoring progress toward goals</td>
<td>Tracking task and progress toward mission accomplishment, interpreting system information in terms of what needs to be accomplished for goal attainment, and transmitting progress to team members</td>
</tr>
<tr>
<td>Systems monitoring</td>
<td>Tracking team resources and environmental conditions as they relate to mission accomplishment, which involves (a) internal systems monitoring (tracking team resources such as personnel, equipment, and other information that is generated and contained within the team) and (b) environmental monitoring (tracking the environmental conditions relevant to the team)</td>
</tr>
<tr>
<td>Team monitoring and backup behavior</td>
<td>Assisting team members to perform their tasks; assistance may occur by (a) providing a teammate verbal feedback or coaching, (b) helping a teammate behaviorally in carrying out actions, or (c) assuming and completing a task for a teammate</td>
</tr>
<tr>
<td>Coordination</td>
<td>The process of orchestrating the sequence and timing of interdependent actions</td>
</tr>
<tr>
<td><strong>Interpersonal processes</strong></td>
<td></td>
</tr>
<tr>
<td>Conflict management</td>
<td>Preemptive conflict management involves establishing conditions to prevent, control, or guide team conflict before it occurs; reactive conflict management involves working through tasks and interpersonal disagreements among team members</td>
</tr>
<tr>
<td>Motivation and confidence building</td>
<td>Generating and preserving a sense of collective confidence, motivation, and task-based cohesion with regard to mission accomplishment</td>
</tr>
<tr>
<td>Affect management</td>
<td>Regulating member emotions during mission accomplishment, including (but not limited to) social cohesion, frustration, and excitement</td>
</tr>
</tbody>
</table>

capture only a cross-section rather than the entire performance episode as it progresses over time. Marks et al. (2001) specifically focused on this temporal aspect of team performance when developing the previously mentioned 10 core teamwork processes. They posited that different team processes are critical during different phases of task execution. Specifically, teams cycle through action and transition phases, and depending on which phase the team is currently in, the processes occurring will differ. Therefore, measuring team performance at one point in time may paint a completely different picture than measuring it at a different point in time or even at multiple points in time. Researchers have long been working to overcome the measurement challenges inherent in the complex, multidimensional, dynamic nature of team performance (e.g., Salas, Priest, & Burke, 2005).

Guzzo and Dickson (1996) reviewed the literature on team performance and noted empirical support for several important team process and outcome variables. Research has found positive associations between cohesiveness and performance (Evans & Dion, 1991; Guzzo & Shea, 1992). For example, Evans and Dion (1991) found that 18% of variance in performance was accounted for by cohesion after correcting for measurement error. Campion, Medsker, and Higgs (1993) looked at several variables, including composition. They found a positive relationship between team size and effectiveness \((r = .23, p < .05)\) and a composite measure of composition including heterogeneity, flexibility, size, and preference for group work \((r = .21, p < .05)\). Although Campion et al. found no significant effect, or a negative one \((r = -.05, ns)\), on heterogeneity and performance, Watson, Kumar, and Michaelsen (1993) found time to be influential, specifically that heterogeneous teams (operationalized by cultural diversity) who worked together long enough overcame the performance deficits relative to homogeneous teams. Others have found positive relationships between performance and (a) familiarity (e.g., Goodman & Leyden, 1991; finding increases from 1.8% to 11% in productivity), (b) leader expectations (e.g., Eden, 1990; \(\omega^2 = .19\) and .17 for performance operationalized as theoretical specialty and practical specialty, respectively), (c) leader mood (e.g., George & Bettenhausen, 1990; \(r = .43, p < .01\), for prosocial behavior, which was significantly correlated with sales performance), (d) motivation as defined by team efficacy and group potency (e.g., Gully, Incalcaterra, Joshi, & Beaubien, 2002; \(p = .41\) and .37, respectively), (e) self-efficacy (e.g., Earley, 1994; effort and self-efficacy accounted for a .66 change in \(R^2\) over demographic variables alone, with all variables together accounting for 69% of the variance in performance), (f) team goals (e.g., Weingart & Weldon, 1991; \(r = .45, p < .05\)) and (g) quality of feedback (e.g., Pritchard, Harrell, DiazGranados, & Guzman, 2008; \(r = .45, p < .01\)).

The variability in these findings suggests the presence of moderating variables. Considering the negative effects that Campion et al. found, these moderating variables could change not only the magnitude of the relationships but also the direction. This is important to note for measurement. Because team performance is complex, if relevant impacting variables are not measured, it may lead to incorrect conclusions regarding team performance.

Salas, Sims, and Burke (2005) took a different approach to team performance with the advancement of a theory called the “Big Five in Teamwork.” Their goal was to present a parsimonious framework, highlighting the essence of teamwork. They theorized that teamwork is composed of five core processes: (a) team leadership, (b) team orientation, (c) mutual performance monitoring, (d) backup behavior, and (e) adaptability. They noted the importance of three additional variables of interest: (a) shared mental models (SMMs), (b) closed-loop communication, and (c) mutual trust. This model incorporates not only core team skill-based competencies but also important affective competencies essential for effective team performance. Current research is focusing on the development of performance measures based on this model (e.g., Wiese et al., 2006).

**Team adaptability.** Burke, Stagl, Salas, Pierce, and Kendall (2006) proposed that a critical skill of any team, and by extension, a critical skill to be measured, is the ability of a team to adapt. Team adaptation is defined as a change in team performance resulting from an identified cue or cue pattern, leading to effective and efficient outcomes for the entire team (Burke et al., 2006). Team adaptability is cru-
cial to organizations, as teams are called on to make decisions and solve problems in complex, dynamic environments with increasing frequency. Adaptation of team performance processes is of great importance to those working to understand effective team performance and how teams should alter in response to rapidly changing task demands.

Burke et al. (2006) developed a cross-level mixed determinants model that integrates several perspectives of organizational theory. This multidisciplinary, multilevel, and multiphasic model is one of the most comprehensive conceptualizations of team adaptation available in the literature. The adaptive cycle comprises four process-oriented phases: (a) situation assessment, (b) plan formulation, (c) plan execution, and (d) team learning, as well as emergent cognitive states that function as both proximal outcomes and inputs throughout the cycle. Operating on the assumption that different team processes are critical at different phases of the team cycle (Marks et al., 2001), Burke et al. posited that plan formulation consists of transition processes and that plan execution consists of action processes, with interpersonal processes occurring in both.

**Team cognition.** The study of individual mental models quickly made the leap to the team level, resulting in the concept of SMMs. SMMs are measured by capturing the amount of “sharedness,” or overlap, between a set of mental models, usually through an aggregation technique. Other team-level cognition constructs that can be measured as a dimension of team performance include transactive memory systems, team situation awareness, and metacognition (e.g., J. R. Austin, 2003; Hinsz, 2004; Prince, Ellis, Brannick, & Salas, 2007; see Table 10.4).

Kraiger and Wenzel (1997) suggested that SMMs can be measured on the basis of three key components: knowledge, behaviors, and attitudes, including perceptions, reactions, and structures. They provided numerous methodologies for measuring these three components of SMM, such as card sorts, structural assessments, and attitude perception surveys. They provided several hypotheses with regard to framework of antecedents, outcomes, and components of SMMs. Kraiger and Wenzel also noted another important issue with regard to measuring SMMs: measure weighting. They postulated that what measures to use and how to weight them are very situation specific yet cautioned that any selected measures should be sensitive to factors such as organizational culture.

### Strategies for Measuring Team Performance

There is no one universally accepted measure of team performance. Guzzo and Dickson (1996) defined team performance effectiveness on the basis of earlier work by Hackman (1987) and Sundstrom, De Meuse, and Futrell (1990). They suggested that team performance effectiveness is characterized by (a) team outputs (e.g., quantity or quality, customer satisfaction), (b) consequences for members, or (c) an increase in ability of a team to effectively perform at a later time. The approaches described below tend to fall under one or more of these three areas.

#### TABLE 10.4

**Components of Team Cognition**

<table>
<thead>
<tr>
<th>Component</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team mental models</td>
<td>Team-level stable mental representations, including key knowledge about undertaking team tasks related to both teamwork and taskwork</td>
<td>Rico et al. (2008, p. 167)</td>
</tr>
<tr>
<td>Transactive memory system</td>
<td>A cognitive system teams use to encode, store, and retrieve information</td>
<td>Lewis (2003)</td>
</tr>
<tr>
<td>Metacognition</td>
<td>What group members know about the ways groups process information</td>
<td>Hinsz (2004, p. 35)</td>
</tr>
<tr>
<td>Team situation awareness</td>
<td>The mental representation associated with a dynamic understanding of the current situation that is developed by team members moment by moment</td>
<td>Rico et al. (2008, p. 167)</td>
</tr>
</tbody>
</table>
Observational approaches. Observational methods of PM, at first glance, are enticing because they offer the objectivity of an outside observer. However, evidence has shown that techniques that use human observers to rate performance have very low inter-rater reliability (Fowlkes, Lane, Salas, Franz, & Oser, 1994). This could be because of several reasons, such as inadequate rater training, leniency effect, or halo effect. Because of this reliability issue, some observational methods have been developed that focus on capturing more objective data such as frequency or existence of specific performance dimensions rather than subjective ratings of behavior as positive or negative. Examples of both rating-based and checklist-based observational methods are described in the following sections. In both categories, measurement requires a third party observer to capture performance.

Behavioral observation scales. The behavioral observation scale (BOS) approach to PM involves the use of observers who provide subjective ratings of the frequency of team performance. In this method, an observer physically watching a team uses a Likert-type scale to rate the amount of times that the team engages in a certain specified process. For example, an observer may be recording the frequency of information exchange in a team. Behaviors representative of information exchange would be rated on a scale from 1 to 5, representing intervals between none and always. Specifically, the item may ask “How often does this team share information about the environment with each other?” The observer would then rate the behavior as happening never, seldom, sometimes, frequently, or always on a 1-to-5 scale.

The distinguishing characteristic of this measurement method is that it assesses the typical behavior of a team over time. However, this characteristic also makes ratings more susceptible to recency effects. Over time, observers may ignore earlier behavior in favor of the more recently observed behaviors. For example, a team may begin their task by sharing information 50 times an hour but may slow down to 20 times an hour as they become more familiar with the task and each other. Because the most recent behavior viewed by the observer was a subjectively “low” 20 utterances an hour, they may rate this team as “seldom” communicating, when on average they were actually communicating 35 times an hour.

Another problem with the BOS method is the subjectivity of the rating scales. Specifically, frequency is a very subjective concept. Some people may consider the same objective amount of communication (e.g., 50 utterances) as never communicating, whereas others may consider that as always communicating. Finally, there is a temporal aspect that bears discussion. Consider the information exchange example above. Perhaps in the beginning, teams should have been communicating 50 times, but as team members become more familiar with each other, they should only communicate 20 times. If the measurement scale is not adapted to reflect optimal levels of performance, teams may inadvertently receive low ratings (i.e., 20 times per hour as low), when in reality that amount of information exchange is quite appropriate for effective performance.

Behaviorally anchored rating scales. The behaviorally anchored rating scales (BARS; Smith & Kendall, 1963) method is very similar to the BOS approach in that it requires an outside expert to observe, classify, and rate behavior (Kendall & Salas, 2004). Originally, this measurement technique was created to evaluate individuals; however, BARS has been readily adopted for use within the team performance arena. Just as in BOS methods, the observer rates the occurrence of team behavior on a numerical scale. However, rather than simply rating the frequency of behavior within the team, behavior is rated in regard to quality. There are specific examples of high-quality and low-quality behavior attached, or anchored, to each rating point in the scale. For example, the scale may range from 1 to 5, with 1 being poor behavior and 5 being excellent behavior. The behavioral anchors are usually generated by SMEs who provide performance episodes that represent both exceptional and unacceptable performance in the specified situation. These behavioral examples, in the form of short written descriptions, are intended to facilitate more accurate ratings by observers by making sure that similar behaviors are rated as the same number across raters. The point of this measurement technique, as used with teams, is to capture the relative frequency of a behavior as well as providing context as to where that frequency lies on a good to bad continuum. Essentially, just noting that a behavior
occurred 20 times does not provide evaluative information as to what level of quality that frequency represents. Therefore, BARS provide more usable information than BOS in most cases.

There are some issues inherent in the BARS method, however. Because behavioral anchors focus on specific types of behaviors, observers tend to watch for those types of behaviors and rate performance on the basis of them regardless of the possibility that the team is using an equally successful, but different, behavior and regardless of the overall performance (Kendall & Salas, 2004). This issue actually relates back to the issue of criterion dimensionality, meaning several different approaches can be used to successfully complete a given task. Therefore, BARS are best suited for PM in work situations that require very specific behavioral responses and in which criterion dimensionality is not a likely possibility.

**Event-based performance measurement.** Event-based measurement techniques, often used in training exercises, are “event-based” because they involve systematically scripting events into a relevant exercise, task, or scenario to trigger specific behavioral responses from the team in question (Fowlkes et al., 1994). In this measurement approach, similarly to BOS and BARS, specific behaviors are identified by SMEs as critical, and these behaviors are recorded as they are observed. It is the high level of control over the appearance of relevant behaviors that makes this a unique measurement technique (Salas, Burke, Fowlkes, & Priest, 2003). The intention of scripting scenarios and identifying behaviors a priori is to increase the level of reliability in measurement. Additionally, event-based measurement is distinct from BOS and BARS techniques because behaviors are not necessarily rated on the basis of quality. One specific method for event-based team PM is known as the targeted acceptable responses to generated events or tasks (TARGETs) methodology (Fowlkes et al., 1994).

**Targeted acceptable responses to generated events or tasks.** In the TARGETs method of PM, a checklist of specific, observable behaviors is generated on the basis of the purpose of the observer and the task characteristics of the team being observed. For example, TARGETs for a helicopter aircrew may include statements such as “Pilots question unsafe navigation procedure” or “Pilots acknowledge communications” (Fowlkes et al., 1994, p. 52). These are specific, easily observable behaviors that an observer could capture during team task execution. Therefore, a trained observer could simply watch the team in action and check off boxes as each behavior occurs. One important characteristic of the TARGETs method is that the events within the scenario being measured are controlled to elicit the behaviors of interest, because routine scenarios generally will not result in a wide enough range of observable behaviors. This makes the TARGETs methodology of PM, or any other event-based approach, much better suited for laboratory-based research than for field or practical use. Additionally, it may be noted that this particular technique (among others used within team PM) can be equally applied to individual PM. It is important to remember that team performance is frequently measured at the individual level and then aggregated to the team level. Therefore, it is important to use measures that fully capture individual performance yet can account for the complexities of team performance.

**Team dimensional training.** Another form of event-based measurement, specifically designed for training teamwork skills, is team dimensional training (TDT; Smith-Jentsch et al., 1997). This training system focuses on measuring and improving four core teamwork behaviors: information exchange, initiative or leadership, supporting behavior, and communication. As in all event-based approaches, a training scenario is carefully designed to provide ample opportunities for the trainees to engage in the four teamwork behaviors. As the trainees go through the scenario, an instructor observes and records examples of strong and weak execution of the teamwork behaviors using a coding sheet. The coding sheet includes a column with the times when each scripted event should occur and a column where the instructor writes a description of how team performs in reaction to those events.

**Communication analysis.** Another common method for examining teamwork performance has been through the analysis of communication transcripts (Salas et al., 2003). The analysis of communi-
cation data generally takes two forms: content analysis and flow analysis. Content analysis focuses on analyzing the linguistic content of the communication data, such as the topics of discussions, the frequency of specific words, or the frequency of questions asked. Typically in content analysis, communication utterances are categorized into groups that represent behavioral constructs such as information exchange or backup behavior, and the frequency of those utterances represent the amount of that behavior occurring. Flow analysis differs from content analysis in that it ignores the meaning or content of the communications and focuses only on the pattern of the team’s interactions. For example, flow analysis might investigate whether there is a consistent pattern in the types of communications occurring, such as responses occurring immediately after questions.

This is a unique method for during-performance PM in that it requires no participation during the team task except for the presence of audio or video recording equipment. The actual evaluation of performance is performed post hoc. Archival measurement such as communication analysis is appealing because the source of the data is permanent, easy to access, and objective in nature. The problem is that measures capture performance well after it has occurred and therefore may draw an outdated picture. Additionally, analysis of communication transcripts requires the prior development of a coding or classification scheme to guide coders in interpreting communication data in the context of teamwork. As this technique requires outside raters to qualitatively assess behavior, it is critical that the raters are properly trained in the classification scheme. Classification schemes for communication analysis can take many forms. Communication transcripts could be coded for the frequency of communication, the content of communication, the pattern communication, or a combination thereof. For example, the number of request for information could be counted, as well as whether these requests for information were regarding the taskwork of others or the environment. These communication instances could be considered as representations of team and systems monitoring, which are two of the critical team processes identified by Marks et al. (2001).

Automated measurement. Automated measurement techniques are one of the more recently discussed methods in team performance literature. This is not a set of specific measurement tools but is rather a specific strategy for implementing various performance metrics. Specifically, automated computer systems can be used to continuously monitor team processes such as communication utterances or body movements (Kendall & Salas, 2004). The recorded behaviors are compared against an expert standard, and the system can provide automated feedback to the team. Automated measures are less obtrusive than other measures (Salas, Priest, & Burke, 2005), but they can be used to measure only overt behaviors and are often quite expensive to implement and maintain.

Summary
The team-level measurement techniques discussed in this section included observational methods such as BOS and BARS, event-based methods such as TARGETS and TDT, as well as communication analysis and automated measurement (see Table 10.5). There are clearly some parallels between team-level and individual-level measurement in that BOS and BARS are both rating methods just as PAs or multiple-source feedback are. However, the team PM literature tends to be heavily focused on the use of measurement for feedback and training purposes, and places a heavy emphasis on the measurement of performance at multiple levels. As we move on to the organization PM literature, the emphasis shifts from multiple levels of analysis to multiple simultaneous dimensions of performance. This point is more fully developed throughout the remainder of this chapter.

PERFORMANCE MEASUREMENT FROM THE ORGANIZATIONAL PERSPECTIVE
The third level of performance measurement focuses on the processes and outcomes of the organization as a whole. In the following sections, we define organizational performance, summarize the key theories of organizational performance, and present the most common measurement strategies for capturing organizational-level phenomena.
Defining Organizational Performance
The larger system under which organizational teams are embedded provides the context for team performance (Guzzo & Dickson, 1996). Researchers have long argued that there has been a lack of emphasis on tying team performance to the overall organizational performance (Levine & Moreland, 1990). McGrath (1991) argued that teams are partially nested and loosely coupled to the broader organization. This refers to the fact that individuals are often part of more than one team and that teams can be part of more than one organization. Team performance, thus, impacts organizational performance.

The literature pertaining to organizational PM is incredibly vast and ranges from theories and reviews of performance (e.g., Folan & Browne, 2005; Herman & Renz, 2008) to case studies and descriptions of specialized PM systems developed for individual companies (e.g., Bhasin, 2007; Khan & Wibisono, 2008). A comprehensive review of this literature is far beyond the scope of this chapter; thus we focus on several of the more commonly studied theories and strategies for organizational PM, as well as some of the newest approaches.

Traditionally, scholars of organizational performance have focused on measuring financial outcomes (Ghalayini & Noble, 1996). Financial outcomes such as return on investment, productivity, or sales per employee provide a concrete, easily accessible measure of overall organizational performance. Yet in terms of assessing the entire organizational performance domain, these measures may be deficient. Therefore, the focus on organizational-level PM has shifted from financial outcomes to more integrated measures of performance that include financial performance along with other dimensions of performance such as customer service and organizational learning. Some define organizational performance from a systems management approach, and therefore collect measures of both internal and external performance information (Jensen & Sage, 2000). Only very recently has organizational performance literature begun to conceptualize performance as a process and explore the measurement of processes in relation to strategic goals and company policy (Nenadal, 2008).

Theories of Organizational Performance
The measure of organizational-level performance used depends on the model of organizational effectiveness

### TABLE 10.5

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral observation scales</td>
<td>A subjective evaluation of team performance completed by an uninvolved third-party observer, usually based on a Likert-scale</td>
<td>Dominick et al. (1997); Kendall and Salas (2004); Taggar and Brown (2001)</td>
</tr>
<tr>
<td>Behaviorally anchored rating scales</td>
<td>A subjective evaluation of team performance completed by an uninvolved third-party observer that includes illustrative examples of good and bad behavior on which to anchor the ratings</td>
<td>Kendall and Salas (2004); Motowidlo and Borman (1977)</td>
</tr>
<tr>
<td>Targeted acceptable responses to generated events or tasks</td>
<td>A measurement system that involves scripting a scenario to elicit behaviors of interest and then completing a checklist of those behaviors</td>
<td>Dwyer et al. (1997); Fowlkes et al. (1994)</td>
</tr>
<tr>
<td>Team dimensional training</td>
<td>A form of event-based measurement that focuses on measuring and providing feedback regarding information exchange, initiative or leadership, supporting behavior, and communication</td>
<td>Smith-Jentsch et al. (1997, 2008)</td>
</tr>
<tr>
<td>Communication analysis</td>
<td>The analysis of communication data or transcripts for indicators of performance behaviors</td>
<td>Bowers et al. (1998); Dong (2005); Landauer et al. (1998); Muniz et al. (1996)</td>
</tr>
<tr>
<td>Automated measurement</td>
<td>Any measurement system that uses automated computer systems to continuously monitor and record team processes such as communication or body movements</td>
<td>Kendall and Salas (2004); Salas, Priest, and Burke (2005)</td>
</tr>
</tbody>
</table>
serving as the theoretical basis. Ahmed (1999) reviewed several of the most widely studied models of organizational effectiveness, including the goal model, system model, internal process model, human relations model, and the political approach. These models differ in their emphasis and prioritization of different dimensions of performance.

The goal model (Georgopoulos & Tannenbaum, 1971) defines organizational effectiveness in terms of an organization’s achievements of its stated official goals. This is one of the earliest and most dominant models of effectiveness, given it directly attempts to align PM with the organizational strategy. The system model of effectiveness views organizations as open systems that work in a close relationship with the external environment. In this model, effectiveness is seen as the organization's ability to survive and adapt in a changing environment. The internal process model conceptualizes effectiveness as the extent to which the internal business processes of an organization are smooth, orderly, continuous, predictable, and with minimal conflict. The human relations approach has a person-centered focus and maintains employee needs (i.e., employee satisfaction, self-efficacy) as the most important aspect of organizational effectiveness. The political approach is a very unusual model of organizational effectiveness, and it uses criteria such as “responsiveness, accountability, representativeness, and adherence to democratic values” (Ahmed, 1999, p. 544). Clearly, each of these theories focuses on very different aspects of organizational performance and would require very different methods of PM.

**Total quality management.** From a business perspective, total quality management (TQM) represents a management philosophy of satisfying the customer’s requirements continually, at a low cost, by involving everyone’s daily commitment (Kanji, 1990). TQM defines quality as a process that must be managed. The objective of TQM is to train all levels of an organization to accept a new set of rules, methods, and lifestyles that focus on continuous improvement (Chung, Tien, Hsieh, & Tsai, 2008). There are four stages to the TQM process: (a) identifying and collecting information about the areas that need improvement in the organization, (b) making sure that management understands and accepts the TQM philosophy, (c) identifying and resolving issues by involving all of management and supervision in a proper scheme of training and communication, and (d) starting new initiatives with new targets and spreading the improvement process to all aspects of the organization including supplier and customer links. TQM has a definite multilevel perspective in that it advocates a philosophy of continuous PM and improvement in all systems within a business. A recent study examined the value of TQM in actual organizations and found that 15 organizations that used TQM had above-average financial ability (Chung et al., 2008).

**Knowledge management.** The knowledge management (KM) approach to PM budded out of the organizational management paradigm shift from focusing on managing physical goods to focusing on managing intangible assets (Nielsen, 2005). In this management philosophy, knowledge is viewed as a critical resource in an organization, and management of this resource can lead to competitive advantage. Of course, to manage knowledge in an organization, it must be measured. The literature dealing with KM can generally be grouped into content and process perspectives (Nielsen, 2005).

The content perspective looks at the “what” of knowledge in organizations, specifically examining the categorization and transferability of different types of knowledge in organizations (Nielsen, 2005). Organizational knowledge can be categorized into explicit and tacit knowledge. **Explicit knowledge** is the knowledge that is exchanged using formal, systematic language and includes things such as explicit facts (Nielsen, 2005). Often, this knowledge can be found in instantiated formats such as books, databases, and computer programs (Small & Sage, 2005/2006). **Tacit knowledge**, on the other hand, is more intuitive, non-verbalized knowledge that is not directly articulated but still exists (Nielsen, 2005). This type of knowledge is hard to put into words and usually is rooted in contextual experiences (Small & Sage, 2005/2006). Other content-based KM research has looked at the holistic concept of organizational knowledge and the distinctions between data, information, and knowledge (Small & Sage, 2005/2006).
The process perspective of KM focuses on the “how” of organizational knowledge and therefore looks at the different stages of information processing in which knowledge is “embodied, embedded, enbrained, encultured, and encoded” within members of the organization (Nielsen, 2005, p. 4). One goal of process perspective KM research is to uncover how the processes for accumulating and internalizing knowledge can be enhanced or improved (Nielsen, 2005).

Strategies for Measuring Organizational Performance
The approaches used to measure organizational performance have changed over the years. Earlier measures included simpler outcomes, such as financial data and balanced scorecards (BSCs). More complex, integrated measures have emerged, such as Six Sigma and ProMES. Each of these approaches is described in more detail in this section.

Financial data. As mentioned previously, traditional organizational PM literature focused on financial outcome measures. Financial measures are easy to obtain, easy to interpret, address the bottom line, and often are already recorded in organizational reports, thereby removing the need for a separate PM system. This makes them an enticing option for organizations looking to measure their overall levels of performance. Productivity, earlier defined as the ratio of effectiveness to the cost of achieving that level of effectiveness, has been one of the most widely used indicators of financial performance in organizations (Ghalayini & Noble, 1996).

Although financial performance measures appear to be a simple, easy-to-use option for measuring organizational performance, the enticing simplicity of financial data brings with it many inherent problems. One of the problems attributed with financial measures such as return on investment is the potential for managers to engage in short-sighted decision making to maximize the short-term return, and consequently, sacrificing the long-term well-being of the firm (Pandey, 2005), a potential contributor to the current economic crisis. Along the same lines, some have argued that financial measures “tell the story of past events” and therefore are inadequate for information age companies moving at a fast pace (Paranjape, Rossiter, & Pantano, 2006, p. 6). Additionally, they may not be the best choice of organizational performance indicators for nonprofit organizations because profit is not the primary focus of those groups.

Ghalayini and Noble (1996) outlined several general limitations of traditional financial measures. One limitation is the missing link between measuring performance in financial terms and the improvement efforts needed to improve that financial performance. It is often difficult to diagnose the causes of performance problems at the individual or team level if all one has is organizational-level financial data. It is this problem of translation between the process of performance and its outcomes that makes financial performance measures less than ideal.

Additionally, traditional financial measures do not take into account the strategy or goals of the organization as an entity, unless those goals are purely financial to begin with. For example, an organization may have high-level qualitative goals such as becoming well known in a particular target population or developing a reputation for timely and accurate service. The attainment of these goals cannot be measured using financial data. This is problematic, especially if the goals are central to the organization’s identity. This is where the measurement of individual- and team-level behaviors, cognitions, and attitudes can improve a PM system.

Balanced scorecard. One of the first PM tools created in response to the limitations of traditional financial measures is known as the BSC. The BSC is “a system of combining financial and nonfinancial measures of performance in one single scorecard” (Pandey, 2005, p. 51). BSC is not a strategy but is rather a management tool focusing on the financial and nonfinancial goals of an organization (Pandey, 2005). It was developed in response to the realization that quite often, nonfinancial goals are the drivers of business success, and therefore purely financial measures of performance are not sufficient alone. This integrated PM system complements financial measures with critical nonfinancial perspectives (Paranjape et al., 2006). A recent estimate...
states that over 60% of organizations use a scorecard (Kaplan & Norton, 2005).

BSC suggests the inclusion of measures of four perspectives—(a) financial, (b) customer, (c) internal business processes, and (d) learning and growth—although it does not have to be restricted to these four dimensions of performance (Pandey, 2005). All four of these perspectives are necessary to get an accurate picture of organizational performance. One of the unique aspects of this measurement approach is that BSC implies a complex causality between the four dimensions, with learning and growth influencing the internal business processes, and the processes have an impact on the financial outcomes, either directly or through the customer perspective (Dror, 2008).

Financial measures such as growth, profit margin, return on investment, and so on, are backward-looking indicators but are necessary to see whether process improvements are translating into financial success. Customer-focused measures include customer satisfaction, customer retention, market share, and customer profitability. For example, poor customer satisfaction is a leading indicator of future performance decline (Pandey, 2005). The internal process perspective looks at the quality of the business processes within the organization, and the key objectives are process improvement and suppliers’ relations. Finally, the learning and growth perspective focuses on innovation, creativity, and capability. Measures of learning and growth include employee satisfaction, employee retention, and employee productivity. In the BSC approach, each of the four perspectives includes (a) objectives to attain (e.g., higher customer satisfaction, higher return on investment), (b) measures of those objectives (e.g., financial data, customer survey data), (c) target values of those measures, and (d) initiatives needed to achieve those targets. The success of BSC depends on the clear identification of nonfinancial and financial variables, their accurate measurement, and linking performance to rewards and penalties (Pandey, 2005).

As literature on BSC has grown, there have been two distinct and conflicting viewpoints emerging. One division of the literature advocates the successes of BSC. The main advantage of the BSC system is the ability to translate an organization’s vision and central mission into tangible, achievable objectives and measures (Kanji, 2002; Pandey, 2005). BSC is unique from other measurement systems in that it contains both the outcome measures (i.e., financial measures) and the performance drivers of those outcomes (i.e., customer, innovation, internal process; Kanji, 2002). Kanji (2002) delineated several additional strengths of the BSC approach. While focusing on multiple dimensions of performance to get a more holistic view of organizational effectiveness, it simultaneously limits the number of measures so as to avoid information overload. It is also a relatively flexible PM approach and can be individualized for each organization. Finally, it places strong emphasis on customers and the market, which are often ignored by traditional measures.

Alternatively, literature also criticizes the approach and calls attention to the lack of scientific evidence linking BSC to improved organizational performance (Paranjape et al., 2006). Other criticisms of BSC claim that it lacks a mechanism for building and maintaining relevance of measures once they are defined (Dror, 2008). In response, Kaplan and Norton (2004) introduced the concept of a strategy map, which provides such a mechanism for connecting strategic objectives with each other. Kanji (2002) described several other weaknesses of BSC. It was designed as a conceptual model and, therefore, is hard to convert into measurement. It is not a comprehensive system approach and leaves out important stakeholders in addition to the role of individual employees and suppliers. The effectiveness of the strategy map has yet to be empirically investigated.

**Six Sigma.** Six Sigma is another performance management technique that has gained increasing popularity. Technically, the name *Six Sigma* can be translated into “3.4 defects per million opportunities” (Banuelas & Antony, 2003). This technique was developed at Motorola as a response to a large loss in productivity due to a lack of quality (Raisinghani, Ette, Pierce, Cannon, & Daripaly, 2005). Although the application of Six Sigma is spreading into increasingly different business sectors, historically, Six Sigma has been predominantly manufacturing-based (McAdam, Hazlett, & Henderson, 2005). This method of performance management “relies on
planned change, team-based collaboration, a focus on performance improvement, a systems perspective, and reliance on the scientific method and statistical methodologies” (Jeffery, 2005, p. 21). Six Sigma is a very process-focused method and therefore attempts to reduce errors and mistakes by experimenting with processes until more stable and robust processes are achieved.

When Six Sigma was first developed, it was primarily seen as a statistical method for tracking and reducing variability in processes (McAdam et al., 2005). However, as the popularity and use of Six Sigma have increased, the approach has grown into a much more complex operations improvement and problem-solving methodology. One of the most interesting aspects of the Six Sigma approach is the change in organizational culture that is often associated with its use. Many organizations using Six Sigma view it not as a statistical method for reducing error, but as a complete business philosophy requiring high levels of organizational and management support. This buy-in is crucial to the success of the strategy.

The Six Sigma method can be implemented using two different approaches—a continuous improvement design (reactive strategy) or a design–redesign approach known as Design for Six Sigma (DFSS; Banuelas & Antony, 2003). The continuous improvement design follows a sequence of five phases (a) define, (b) measure, (c) analyze, (d) improve, and (e) control (DMAIC). In the define phase, the problem to be addressed is identified and defined, and the critical stakeholders are identified. In the measure phase, the measurement capability of the organization is assured, current levels of performance are identified (most often using existing corporate performance measures), and goals for improvement are then set. In the analyze phase, the causes of problems and the key variables that may be linked to defects are identified. In the improve phase, experimentation is used to quantify the impact of variables on the process, and the process is subsequently improved. Last, in the control phase, continuous monitoring and adjustments are used to maintain the improved process (Goh & Xie, 2004). This improvement approach assumes the initial process is essentially correct and thus only needs to be adjusted for optimal performance (Banuelas & Antony, 2003).

Raisinghani et al. (2005) described five general steps in the DMAIC approach. These include (a) determining the appropriateness of the existing PM equipment through analysis, (b) looking for process deviations that require improvement, (c) using the “design of experiments” technique, (d) conducting a failure mode and effects analysis, and (e) taking one final measure of quality to ensure that Six Sigma levels have been achieved.

DFSS follows a slightly different sequence of phases: (a) define, (b) measure, (c) analyze, (d) design, and (e) verify (DMADV). It differs only in that rather than improving an existing process, and controlling that change, a brand new process is designed and then verified. DFSS is a more proactive approach in that it involves designing processes capable of reaching Six Sigma levels, effectively preventing problems before they arise rather than after (Banuelas & Antony, 2003). This approach does not assume that the initial process is correct; instead, it aims to replace the process with a better, more effectively designed process. Therefore, both efficiency and effectiveness of the process can be improved using the DMADV approach.

As with the implementation of any PM system, the context surrounding the Six Sigma approach is critical. McAdam et al. (2005) delineated a list of key steps in the Six Sigma processes on the basis of a review of the literature. The organization must gain senior manager support and involvement to create the desired deep level. A team of specialists must be trained to lead the process, requiring an initial investment of time devoted to training and developing a set of experts in the process. Much of the limited empirical evaluation of Six Sigma focuses on case studies (e.g., Antony, Kumar, & Madu, 2005).

Productivity Measurement and Enhancement System. ProMES is an intervention designed to improve performance by actively measuring it and using those measures to provide detailed performance-based feedback (Pritchard, 1990; Pritchard et al., 2008; Pritchard, Holling, Lammers, & Clark, 2002). A typical ProMES intervention involves a design team (composed of people responsible for doing the work) who identifies the main objectives of the unit and develops quantitative
measures of these objectives. These measures are designed to be valid measures of all the important aspects of work required from the unit. The system is then reviewed and approved by higher management. On the basis of the measures, a regularly occurring feedback report is provided to the work units to guide desired improvements.

The conceptual framework underlying ProMES is the NPI theory (Naylor, Pritchard, & Ilgen, 1980). NPI theory is a refinement of expectancy theory (Arvey, 1972; Campbell & Pritchard, 1976; Vroom, 1964) with a novel approach to motivation. Motivation within NPI is considered a process of allocating energy and time to various activities (Pritchard et al., 2002). The NPI motivational process involves acts (individual behavior), focusing on both direction (why certain acts are chosen over others) and amplitude (the amount of effort devoted to completion of the act). These acts lead to the generation of products (individual outputs). These products are then evaluated, which results in evaluations. These evaluations come from supervisors, peers, self, subordinates, family, and so on. Outcomes are tied to those evaluations and can either be intrinsic (i.e., increased satisfaction with one’s work) or extrinsic (i.e., pay raise, promotion). Outcomes are motivating because of their tie to needs satisfaction, which are very individual desires regarding everything from status to health and more. Acts, products, evaluations, outcomes, and needs satisfaction are combined into motivational force, which refers to an individual belief that changes in devoted time and energy (effort) toward different acts (tasks) will lead to changes in the amount of needs that are satisfied (Pritchard et al., 2002).

A key characteristic of NPI is contingencies. Between each of the motivational components are component links (contingencies), signifying the interconnectivity among components (Pritchard et al., 2002). For example, actions generate products, thus the creation of products is contingent on the amount of effort directed toward acts or behaviors that directly relate to generating the product. Each contingency is based on individually perceived relationships. Again looking to the actions to products link, acts-to-products contingencies depict the relationship between the amount of effort or time devoted toward an act and the expected amount of generated product.

ProMES consists of seven steps (Pritchard, 1990; Pritchard et al., 2002). For a more detailed review, see Pritchard (1990). The first step is creating a design team of seven to eight people who are ultimately responsible for creating the new system. On the basis of the notion of participation in decision making, the team is composed of employees who do the work under evaluation. The second step focuses on the identification of overall unit objectives, developed through group discussion to consensus. The objectives should describe the purpose of the work group. All dimensions of the work must be reflected in the overall objectives. The third step involves identifying indicators or quantifiable measures of effectiveness to determine how well the objectives are being met. The objectives and indicators are presented to upper management for approval, and any disagreements are discussed to consensus. The fourth step moves toward defining contingencies. Contingencies are basically a function of the amount of the indicator as compared with its value to the organization. Indicators are charts, plotting effectiveness on the y-axis and the objective on the x-axis. The fifth step is to design the feedback system. Data are collected from each indicator, and indicator effectiveness scores are calculated. These individual scores are aggregated into an overall effectiveness score. A feedback report is then prepared with these data and is compared with historical data to determine improvements and well as areas that require further attention. The sixth step is the actual feedback session between the supervisor and employee, who discuss causes for improvements and any noted decreases, and devise a plan for continued improvement or changes. The seventh step involves monitoring the project overtime. If it is determined after several feedback sessions that some aspect of the measurement system needs to be changed, then the entire process should start again to come to agreement on how it should be fixed. If there are significant changes in the work or policy changes, then the system should also be reviewed.

This system is highly effective. Not only does it lead to large productivity improvements in many different types of settings, these effects have also been
shown to last over time (Pritchard et al., 2008). The mean weighted effect size is 1.44, which is equivalent to the 93rd percentile of performance under baseline. Pritchard et al. (2008) also investigated a number of potential moderators. They found that the degree to which the study followed the ProMES methodology, the quality of the feedback given, whether changes were made to the new system, the degree of interdependence of the work group, and the degree of centralization of the organizations moderated the effectiveness of a ProMES intervention.

Summary

Financial data, BSC, Six Sigma, and ProMES represent some of the most widely studied and discussed approaches for measuring organizational performance (see Table 10.6). As the science of organizational PM has grown, there has been a definite push toward more integrated, multidimensional PM systems. Financial data alone are no longer considered an appropriate gauge of organizational effectiveness. There has also been a push toward PM systems that are more diagnostic of the causes underlying performance. We contend that taking a multilevel approach to PM is one of the most effective ways to diagnose organizational performance. This approach is described in detail in the following section.

### Table 10.6

#### Organizational-Level Measurement Techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial data</td>
<td>Outcomes measures focused on financial success such as return on investment or productivity</td>
<td>Ghalayini and Noble, (1996); Pandey (2005); Paranjape et al. (2006)</td>
</tr>
<tr>
<td>Balanced scorecard</td>
<td>A performance measurement system that combines financial and nonfinancial measures into a single scorecard for organizational effectiveness</td>
<td>Ahn (2005); Kanji (2002); Kaplan and Norton (1996); Norreklit (2000); Pandey (2005); Paranjape et al. (2006); Schwartz (2005)</td>
</tr>
<tr>
<td>Six Sigma</td>
<td>A performance measurement system that focuses on improving process design and reducing process error and waste</td>
<td>Antony (2006); Banuelas and Antony (2003); Goeke and Offodile (2005); Goh and Xie (2004); Raisinghani et al. (2005)</td>
</tr>
</tbody>
</table>
again be meaningless. Therefore, it is imperative that organizations have both sound PM strategies as well as sound evaluation strategies for interpreting performance data.

The third step after describing and evaluating performance is to diagnose the causes of effective and ineffective performance. This step is perhaps the most important in the PM process because without an understanding of the causes underlying performance, the development of feedback and training becomes difficult and suboptimal, if not impossible. Identifying deficiencies in performance is not very useful unless one can also identify the underlying causal mechanisms to change to rectify those deficiencies. Diagnosing the underlying causes of effective and ineffective performance is the only way PM can be used to manage and improve performance. This concept of linking specific performance outcomes to specific causal mechanisms to achieve effective PM has been called to our attention by other organizational researchers (e.g., Ittner & Larcker, 2003).

Therefore, we suggest that for any PM system to provide the most accurate description, evaluation, and diagnosis of performance, it should take an integrative, multilevel approach to measurement (see Figure 10.2). This position has been also been advocated quite strongly in the team PM literature (e.g., Salas et al., 2003). Individuals are the units that make up teams, and therefore team PM should include both team-level and individual-level indices. Following the same pattern, organizations are made up of teams, and teams are made up of individuals, so organizational PM should include all three levels of measurement if it is intended to capture the most beneficial and applicable information. The team PM literature focuses heavily on the distinctions between process and outcome, and how measurement of both is necessary if one is to accurately diagnose the underlying mechanisms influencing performance outcomes. Applying this concept to the organizational-, individual- and team-level processes are often the underlying causal mechanisms behind organizational outcomes. Therefore, for organizational PM to be diagnostic, it must capture and connect performance at all active levels of analysis.

Although the paradigm shift in organizational PM from financial measures to more integrated multidimensional measurement systems illustrates a much needed change, organizational performance measures still tend to focus heavily on tangible outcomes at a very high level. Therefore, there is still some level of disconnect between effective performance at the organizational level and the processes that enable that performance at the individual and team levels. This disconnect makes it difficult to translate organizational-level measures into actionable goals for improvement. For example, a corporation may measure profit per unit production to assess overall organizational-level performance and find they are making $5 in profit for each unit produced, and decide they would ultimately like an increase of $2 in profit. Yet, how does this translate into terms of improvement? One can not tell employees to work “$2 harder.” How much extra effort is necessary for two more dollars of profit? What changes in the production process are necessary? What factors influence level of profit? This is why organizational-level PM is important, but not sufficient, for improved organizational performance.

To truly diagnose and improve organizational performance, multiple levels of performance must be captured to link overall financial outcomes to underlying individual- and team-level processes. Effective PM systems should focus on making con-
connections between the different levels of performance and should take advantage of the various techniques available from each perspective. This will allow the organization to gain an understanding of what individual-level behaviors and processes impact team processes and team outcomes, what team-level processes impact organizational outcomes, and so on. By doing this, changes in organizational-level financial performance can be stimulated by managing performance at the employee and team levels. Ultimately, a multilevel understanding of performance will enable the organization to pinpoint the causes behind performance problems and effectively address them.

To illustrate the potential use of multilevel PM, a hypothetical example is described. Assume an organization is planning to implement a new PM system throughout all levels of its personnel. Perhaps the company employs factory workers that work individually to make products on an assembly line, as well as new product development teams that create new ideas for the products, among many other positions. For the sake of parsimony, we discuss only these two specific job positions. Now assume the organization chooses to use an automated measure of performance for the assembly line workers after carefully considering the purpose, content, timing, and setting desired. This automated system records the number of units produced, number of errors made, and several other measures of productivity. This choice of measurement allows for continuous on-the-job, real-time capturing of data. The organization then decides, after careful consideration, that the new product development will be individually rated using 360-degree feedback from themselves, their teammates, and the team leader on several dimensions including innovation, cooperation, and communication. The team will also be rated as a unit on the same dimensions. These subjective rating measures are better suited for measurement of highly creative and innovative jobs. Finally, organizational performance will be measured using overall customer satisfaction surveys as well as archival data reporting the annual profit of the company.

Now, because the organization collected PM data from all three levels of analysis, the organization is capable of linking outcomes to the underlying causal processes. Specifically, perhaps analysis of the data suggests that higher ratings on communication and innovation at the team level within the new product development teams are related to higher levels of customer satisfaction, whereas the number of units produced by assembly line workers is quite unrelated to customer satisfaction scores. This information would allow the organization to pinpoint the process that needs remediation (i.e., communication and innovation within teams) if an increase in customer satisfaction is their top priority. Similarly, they may find that the number of errors made on the assembly line is highly related to overall profit, and therefore an increase in profit would more likely occur because of a reduction in these errors. The main point of this example is that multilevel PM leads to a richer understanding of performance and the links between different processes and different outcomes, and therefore enables the organization to more effectively and efficiently manage their performance.

The connections between different levels of analysis occur not only in a bottom-up direction but also from the top down. Because individuals and teams operate within the particular context of the organization, factors from the organizational level may impact both team and individual functioning (Salas et al., 2003). For example, a lack of resources provided by the organization or a set of organizational procedures may limit the ability of a team or individual to perform in a certain way. Thus it is critical to measure contextual variables at the organizational level to take those into consideration when trying to assess individual or team performance.

EMERGING ISSUES FOR FUTURE RESEARCH IN PERFORMANCE MEASUREMENT

An astounding amount of research has examined PM at the individual, team, and organizational levels. So much so that only a fraction of that research was touched on in this chapter. However, despite the abundance of research in the field of PM, there is always room for growth and further exploration. In particular, as the world changes and technologies advance, several areas have become paramount to the understanding of performance at any level: dis-
ttribution, diversity, dysfunction, and culture. We posit that these concepts represent the future of PM research.

Performance Measurement in Distributed Environments

Globalization has impacted the way organizations work. Corporate offices are distributed throughout the world. Teams are distributed throughout various offices. Yet, another important phenomenon is also occurring—an increase in the number of workers who are working outside of the traditional office environment. A survey of senior-level executives worldwide indicated that two thirds of the global workforce was involved in distributed work arrangements (AT&T, 2004). In the United States, telecommuting, the most common form of distributed work, has increased exponentially over the past several years. In 1997, 11.6 million full-time employees worked remotely at least 1 day per month (WorldatWork, 2006). This number more than quadrupled to 45 million by 2006, the majority of whom are women, ages 35–44 (Gajendran & Harrison, 2007; Roitz, Allenby, & Atkyns, 2002).

The increased prevalence creates a measurement issue—how does one measure the performance of employees when one never observes them actually “performing”? Many theories have put forth the notion that technology-mediated interaction leads to decrements (e.g., Media Richness Theory; Daft & Lengel, 1986). This is seemingly evidence that perhaps distributed performance should be measured differently than performance that is directly observed, to account for these issues. However, others speculate that good measurement is good measurement regardless of the setting and that, for example, what constitutes good PM for teams that are collocated is the same for PM for teams that are distributed—specifically, that similar principles should apply to the development of a system for collocated or distributed employees (Blackburn, Furst, & Rosen, 2003). Blackburn et al. (2003) suggested that the what, how, why and when of PM are closely related in either face-to-face or distributed environments. They outlined several characteristics that should apply to either situation (e.g., performance measures should always be tied to desired organizational goals, the purpose of the measure should be clear). However, there are some noted differences. In face-to-face environments, performance ratings are based not only on the outcomes but also on the perceived effort on the part of the employee. However, this is not possible when measuring virtual performance and thus should be taken into consideration. Additionally, it is important to consider the multiple ways at which to arrive at the same outcome (criterion dimensionality).

These issues and many more need to be systematically addressed by future research. Researchers need to understand the impact of distribution on performance ratings; the value added by distributed team or organizational members; and if performance should be measured differently, then how? Some initial work has been done in this area (e.g., Hofmann, Klar, Mohr, Quick, & Siegle, 1994); however, more research is needed, given the changing nature of technology and the increased prevalence of this type of work. The implications of this research extend beyond teleworkers to any employees who are working away from their home office (e.g., expatriates).

Performance Measurement in Diverse Settings

Today’s workforce is more diverse than ever before. This poses a unique challenge for designing process-oriented PM systems that evaluate fairly across demographic groups. Specifically, criterion dimensionality issues can play a big role in diverse settings. As mentioned previously, criterion dimensionality is the idea that two people on the same job may be equally effective but engage in very different behaviors to reach that level of effectiveness (Borman, 1991). This can become a problem if there are gender or ethnicity differences in the strategies used to complete a task and the measurement system assumes only one correct strategy. For example, men and women may differ systematically on leadership styles because of their differences (i.e., women tend to exhibit more democratic or participative styles whereas men tend to portray autocratic or directive styles; Eagly & Johnson, 1990). If a measure of performance assumes one style of leadership to be more effective than another without actually validating this effect, the measurement system could lead to unfair test bias and adverse impact issues. Therefore, it is critical that researchers...
examine issues of demographic difference in process and how this influences subsequent effectiveness. If findings indeed suggest that multiple strategies can lead to the same level of overall effectiveness, this dimensionality should be reflected in the measurement system.

Performance Measurement and Dysfunction
The study of performance has been pervasive in the industrial and organizational psychology literature for decades. Yet only recently has the field begun to consider unintended consequences of measuring performance. As noted by Grizzle (2002), “we expect that measuring efficiency leads to greater efficiency and measuring outcomes leads to better outcomes, but we don’t always get the results we expect” (p. 363). She noted several examples such as distorting financial data as an unintended consequence of measuring quarterly earnings per share. Many companies, such as Enron, resorted to an incomplete disclosure of financial data in an attempt to manipulate this measure. In terms of process measures, she noted that measuring quantity can lead to a diminished focus on quality. The measure may be indicating positive findings (i.e., high levels of quantity); however, the actual product may be of very poor quality. Others have noted the unintended consequences of measuring performance on the employees. For example, performance monitoring in call centers has been shown to cause employees distress, which impacts their well-being and satisfaction (Holman, Chissick, & Totterdell, 2002).

More research is needed that focuses on these unintended consequences: What are they? Under what circumstances do they occur? How much and in what ways do they affect performance measures? Finally, what types of interventions alleviate the potentially harmful effects of these unintended consequences of PM?

Performance Measurement and Culture
One emerging issue in PM research is the impact of cultural differences on PM. As the globalization of business increases, organizations are beginning to measure performance across national and cultural boundaries like never before. Cultural background influences the qualities and traits people value, and therefore changes the way people rate performance dimensions. For example, Nonaka and Takeuchi (1995) argued that Westerners and Japanese view knowledge differently, which could influence how KM is viewed in multinational organizations. For example, Japanese view knowledge as being primarily tacit, whereas Western cultures tend to focus on explicit knowledge. Therefore, in a PM system that focuses on capturing levels of explicit knowledge, there may be cultural differences between Western and Eastern employees. These cultural differences in how people perceive performance dimensions could indicate that an employee being rated by a manager from one culture may receive a completely different rating from a manager of another culture. The challenge in this situation is either selecting dimensions of performance that are comparable across cultures or building a common understanding of the performance dimensions already being measured. Otherwise, ratings will not be equivalent.

For example, Gillespie (2005) compared 360-degree feedback ratings across Great Britain, Hong Kong, Japan, and the United States and found that there were differences between countries in responses to the survey. Specifically, Gillespie found that respondents from different countries had different understandings of the constructs included on the survey and how these constructs related to each other. Therefore, they responded to the survey differently, meaning that the results of 360-degree feedback across cultures were not comparable. This calls into question the appropriateness of using 360-degree feedback in multicultural contexts because the ratings given by members of different cultures may not be compatible for aggregation. This is a very important issue for multinational corporations to consider, and further research is needed to explore this issue and possible ways to account for these differences.

CONCLUDING REMARKS
The purpose of this chapter was to highlight PM issues as they relate to various levels (individuals, teams, and organizations). Given the scope of the literature on PM developed over the past several
decades, we have presented a brief overview of this topic. By noting issues relevant to individuals, teams, and organizations as well as presenting several measurement strategies common to each level, we hope to synthesize existing literature on this topic in a systematic manner. Although the literature has taught researchers a great deal about PM, there is still much the field does not know. Future research suggestions are designed to spur further investigation in areas that are currently underdeveloped in terms of both scientific and practical understanding. It is clear that progress has been made with regard to improved measurement techniques, the consideration of multiple levels of performance, and a more concerted effort to consider conceptual criteria. Yet for a more broad understanding of PM, more research is still required.

References


Smith, P. C. (1976). Behavior, results, and organizational effectiveness: The problem of criteria. In M. D. Dunnette (Ed.), *Handbook of industrial and
organizational psychology (pp. 745–775). Chicago: Rand McNally.


