Chapter Twelve

Changing Conceptions and Practices in Performance Appraisal

Jerry W. Hedge, Walter C. Borman

As the world of work is transformed to reflect changes in technologies, in the society, and in the demographics of the workforce, our opinion is that the need for evaluating job performance will remain constant. However, the way we go about the process of gathering accurate performance data will require new thinking about old approaches and innovative ideas about new approaches. Based on some current and projected trends in the nature of work, we speculate in this chapter on some of the issues and answers for performance evaluation in the future. First we present a brief overview of these trends.

Everywhere we turn we hear that corporate America is in the process of reinventing itself for the twenty-first century. A fundamental economic change is afoot: the replacement of industry’s century-old electromechanical base by the computer-based service infrastructure, the replacement of the tradesman by the technician. Constant technological advances are changing work patterns, making flexible hours and working from home more commonplace. Workers will be expected to have a variety of skills, be subjected to frequent retraining, and be more self-directed. Management will change, with new priorities of product quality and customer satisfaction reflecting a shift from manufacturing to service (Goldstein & Gilliam, 1990). In the future, flexibility,
in addition to the traditional skills and abilities, will be a hallmark of employment (Lewthwaite, 1993).

We believe that several present and likely future trends in organizations and in the U.S. economy, many of them described in other chapters of this book, will lead to significant changes in the way performance appraisal is accomplished. The following trends seem especially important for influencing future performance appraisal research and practice: (1) the increasing pace of technological change in many jobs and organizations in the United States and elsewhere (Gattiker, 1990); (2) as a consequence of technological change, the accelerating need for worker training and retraining, with the accompanying need to evaluate success in training (London & Bassman, 1989); (3) the movement revolving around worker empowerment, partnerships between management and employees, the encouragement of self-management, and stewardship (Block, 1993; "Empowering Workers," 1991); (4) alternative formats to the traditional office, including working at home (Calem, 1993), telecommuting (Leyden, 1993), "virtual organizations" (Offermann & Gowing, 1990), and other variants; (5) the rapidly increasing use of temporary workers (Diesenhouse, 1993), even at highly skilled and professional levels (for example, "elite temps"); and (6) the burgeoning amount of attention paid to customer satisfaction, with the continuing shift from a manufacturing to a service economy (Peters & Waterman, 1982; Peters & Austin, 1985).

We turn now to a more specific look at performance appraisal, recognizing that how we work and where we work are becoming more diversified than ever before. Within this context, we examine the usefulness of different performance rating sources, including supervisor, peer, self, subordinate, and customer. We also discuss the potential use of other methods of evaluation, such as electronic performance monitoring and work sample testing. Because of the increased reliance on work teams, we examine the concept of team performance and productivity as well. We also discuss the need for a more thorough coverage of the criterion space (for example, prosocial and contextual performance), and how these continued and rapid changes in the workplace are affecting workers' attitudes about, and reactions to, performance appraisal.

Performance Ratings

Ideally, ratings can provide performance scores that are free from contamination and deficiency—that is, perfectly relevant. This is because, first, as long as performance requirements for a job can be articulated and defined, rating scales can be developed to reflect those performance requirements. Second, the rater or raters using these dimensions to make evaluations can, again ideally, average performance levels observed over time and in different job situations to arrive at a rating on each dimension.

This assessment of ratings reflects the potential the method has for yielding accurate performance scores. Unfortunately, rater errors, biases, and other inaccuracies detract from the scores’ validity. Still, it is safe to assume that performance ratings will continue to retain their predominance, if for no other reasons than ease of acquisition and cost, as has always been the case. However, the usefulness or accuracy of a particular rating source may vary considerably according to the structure of the work environment. For example, it may be perfectly appropriate to ask customers to rate the performance of a retail sales clerk, but inappropriate to rely on peer ratings for a telecommuter who spends 75 percent of the work week at home.

In addition, the purpose for which performance ratings are intended appears to play a role in their usefulness. Management uses evaluations for general personnel decisions, such as promotions and terminations. Evaluations also identify training and development needs, pinpointing employee skills and competencies that are currently inadequate but for which programs can be developed. Finally, performance evaluations can be used as a criterion against which selection and development programs are validated. As we discuss the usefulness of different appraisal methods, we do so with an eye to purpose.

The usual practice in most organizations is for an employee to be evaluated by an immediate supervisor. In recent years some organizations have come to realize that other sources besides an employee’s boss can provide appraisals. Because of the dramatic changes affecting the workplace, problems of relying on any one source of appraisal alone may become amplified. The use of
multiple sources for performance ratings has continued to gain acceptance over the last decade and numerous advantages have been cited. It is even becoming fashionable to talk about 360-degree feedback, where evaluations are sought from supervisors, peers, and even subordinates and customers in the belief that feedback from these individuals is enriched by including multiple sources of input. In organizations where 360-degree feedback has been adopted, its primary focus has been developmental, although it has been used for administrative and validation purposes as well (Tornow, 1993a).

While the notion of ratings from multiple sources is not new, the initiative for a formalized program of feedback to managers from their constituents is in its infancy. Much still needs to be done to make such programs viable (see London & Beatty, 1993; Moses, Hollenbeck, & Sorcher, 1993; Tornow, 1993a).

We recommend reliance on more than one source of ratings whenever possible. As the workplace continues to change, we need to be flexible and innovative by using different sources, alone or in combination, as the work environment and purpose of appraisal dictate. We turn now to a closer look at each rating source in terms of its usefulness in the future world of work.

Supervisory Ratings

Changes in the technology of work are likely to have a significant impact on supervisory performance appraisal. In particular, computer-based technologies often allow employees to work with minimal supervision, sometimes at home. In situations such as telecommuting, or even where semi-autonomous work groups are in place, the supervisor will not be in a good position to observe the worker's performance on a regular basis, and other methods of evaluation may be needed.

In other cases the structure of the organization may facilitate close supervision, but because of rapidly changing technologies, supervisors may be less familiar with the technology than are their subordinates. Consequently, technology may undermine the role of the supervisor in performance appraisal (Murphy & Cleveland, 1991). As a result, the supervisor may not be perceived as a competent judge of the technical aspects of subordinates' performance. In such situations, other sources (such as peers) or methods of appraisal (for example, work sample tests using expert evaluators) must be relied on to provide information about technical performance to supplement the supervisory evaluation.

Still, it seems likely to us that the supervisor will remain the primary source of performance appraisal in the workplace of the future. Companies will not abandon supervisor appraisals merely because appraisal accuracy may be affected. Regardless of whether the purpose of appraisal is validation, administrative decision making, or employee development, supervisors will likely retain the responsibility of maintaining close enough contact with subordinates that sufficient information about some technical, motivational, and interpersonal aspects of performance can be gathered as a basis for evaluation. Consequently, research must continue to explore ways to enhance supervisory appraisal accuracy.

Peer Ratings

Peer appraisals have consistently been shown to be reliable. Because multiple peer ratings tend to be collected for each person evaluated, such assessments frequently provide a stable measure relatively free of the bias and idiosyncrasies of a single rater. Also, Latham (1986) has argued that reliability is affected positively by the daily interactions among peers. Not only do peers see how the employee interacts with them, but they also observe how that employee interacts with subordinates and supervisors. Thus peers often have access to more job-relevant information on which to base an evaluation than do other sources.

In terms of the future, peer ratings may be out of the question for employees who work at home, especially those who work in isolation. However, it is conceivable that telecommuters may be linked together and communicate by telephones or computers. In such cases, peers could have access to some aspects of their co-workers' behavior, or the products of that behavior, and thereby provide a relevant evaluation of performance.

The usefulness of peer evaluations may be greater in a team-oriented workplace. Because work groups, to varying degrees, tend to monitor and manage their own performance processes, peer evaluations would seem to be especially relevant with them. Group
members work in close proximity, are able to observe each other’s technical and interpersonal job behaviors, and should therefore be in an excellent position to rate their co-workers. Unfortunately, McEvoy and Buller (1987) noted that peer ratings are not well accepted by raters or those rated except when they are used for developmental purposes. The problem here focuses on asking group members to play a major role in the administrative decision process. Appraising and rewarding individual performance using a peer-rating system may hinder coordination and increase intragroup conflict.

Thus in spite of certain attractive features of peer ratings in a team environment, caution may be advised in the administrative use of performance ratings from peers. The problem associated with acceptance of peer ratings by the peers themselves should be less onerous when the focus is either developmental or for validation/research purposes, or in situations where the team concept is not an integral part of the organizational structure.

Self-Ratings

Much has been written about self-appraisals. Researchers have suggested that self-ratings are unreliable, biased, and inaccurate compared to other ratings sources or more “objective” criteria (for example, De Nisi & Shaw, 1977; Levine, Flory, & Ash, 1977; Nilsen & Campbell, 1993). Other researchers paint a more favorable picture of self-ratings (for example, Carroll & Schneier, 1982; Latham & Wexley, 1981; Shrauger & Osberg, 1981).

Much of the empirical research cited suggests that self-ratings differ from other types of ratings; a major question is whether different means less accurate. Dunnette (1993) concluded that self-descriptions do, indeed, possess accurate components, and suggested that it is premature to presume that others’ ratings should always be used to assess the validity of self-ratings.

Self-ratings have certain advantages. Latham and Wexley (1981) have noted that, from a developmental perspective, self-ratings force the employee to focus on what’s expected in the job. Also, they allow the supervisor to see how subordinates perceive their level of effectiveness. Thus self-appraisals may help clarify and even resolve differences of opinion between the subordinate and boss concerning performance requirements. In addition, self-appraisal helps employees to think about strengths and weaknesses in their performance and, ideally, stimulates activity toward self-improvement. In fact, the recent research on procedural justice suggests that giving the employee an opportunity to provide performance information to the supervisor should also increase the employee’s perceptions of evaluation fairness (Greenberg, 1986).

Self-appraisals may be especially appropriate for employees working in isolation, a scenario that will occur more often as we look toward a future of increased telecommunication. In such work environments, what are the options for assessing performance? Employees working in isolation may know more about their own performance than anyone else and should at least be capable of providing information about their own strengths and weaknesses (Campbell & Lee, 1988).

When performance appraisal is being done for administrative purposes and work is relatively independent, self-ratings may be a necessary component of the appraisal, along with other sources of information. For example, later in the chapter we discuss the use of electronic performance monitoring, a method that can capture product or process aspects of employee performance. An employee working at home could be monitored by computer, and that information along with a self-appraisal could be used to evaluate the individual’s performance. In such instances, if the employees know that self-ratings make up only part of the evaluation, the presence of other performance data may increase the chances that the self-ratings will be accurate.

Subordinate Ratings

Despite evidence that use of subordinate appraisal is on the increase, there is almost no empirical research to support upward appraisal for any purpose. A recent study by Bernardin, Dahmus, and Redmon (1993) does provide reaction data showing that supervisors were supportive of subordinate appraisal as a useful source of data, except when used as a basis for determining pay.

A number of years ago, Latham and Wexley (1981) suggested that subordinate ratings can aid management in identifying supervisors who are promotable because of their skill in managing people. More recently, Tornow (1993b) suggested that subordinates can be seen as providing the most direct source of feedback regarding the target manager’s leadership behavior.
Despite the lack of supportive research on subordinate appraisal, a growing number of companies now use this source for appraising supervisors and managers; these include Amoco, Bank of America, Du Pont, Wells Fargo, Exxon, Tenneco, the World Bank, Johnson & Johnson, and GTE (Bernardin, Dahmus, & Redmon, 1993; Michaels, 1991). Such appraisals are seen as valuable because they can foster team-building, force supervisors and subordinates to gain a better understanding of each other’s jobs, and supply information about a manager’s “people” skills from the subordinate’s point of view.

If organizations of the future increasingly encourage team-based environments, as recent trends suggest (for example, Tannenbaum, Beard, & Salas, 1992), subordinates can contribute useful performance data to the appraisal process. Subordinate appraisals can support team-building by challenging workers and their bosses to discuss and resolve problems of mutual concern. Also, when subordinates must evaluate their bosses on a variety of work performance dimensions, it encourages them to view problems through the eyes of their supervisor, thus allowing them to gain a fuller appreciation of the supervisor’s job. Equally important, once the feedback is received, the supervisor can begin to see concerns from the perspective of subordinates.

Probably the biggest potential problem with subordinate appraisals is the fear that supervisors will exact retribution if an honest but unfavorable appraisal is received. Thus it is critical that organizations maintain anonymity when using such ratings. Unfortunately, ensuring anonymity can become problematic if the work group is small.

At the very least, all should recognize that subordinates may provide valuable information about some aspects of the supervisor’s performance but probably not all dimensions because of their somewhat limited perspective. Obviously, much additional research is needed in this area.

**Customer Appraisals**

Some organizations elicit performance appraisals from persons outside the immediate work environment. Customers are in a unique position to judge what the performance expectations are and how well they have been met, as this reflects customer satisfaction (Tornow, 1993a). Unfortunately, very little research is available to support or refute the use of these types of ratings. A study by Bernardin (1992) provided some support for customer-based appraisals as a source of added and unique information beyond that provided by top-down appraisal.

Ratings by customers or clients are probably the least used or studied source, but with the continued move toward a service economy, such evaluations make good sense. Customers provide another valuable perspective to be used in conjunction with other means of evaluating performance. Most service organizations acknowledge that customer satisfaction is critical for the survival of their business, yet few of them take formal steps to gather information about it. In some ways customer ratings may not face many of the problems associated with accuracy of other rating sources. If any source of appraisal would be brutally honest in its ratings, it is probably customers! Such ratings should be limited to customer service dimensions, however, as customers do not see all aspects of an employee’s job performance. Customer ratings can be used for developmental, administrative, or validation purposes. Again, much additional research is needed in this area.

**Electronic Performance Monitoring**

A relatively new method of gathering performance data about individuals in the workplace is known as electronic performance monitoring. It is already gaining in popularity; according to a 1988 report to the U.S. Congress by the Office of Technology Assessment, approximately 10 million U.S. workers were subjected to secret electronic monitoring that year. Rather than being observed directly by supervisors, employees can be monitored electronically in a variety of ways. Probably the first kind used in industry was telephone monitoring, whereby supervisors (sitting next to the employee, or in their own office) listened in on employee telephone calls. Alternatively, rather than listening to the conversation at the time it occurs, an audiotape of a worker’s telephone call can be made for later analysis. Video cameras can also be used to capture employee activities on tape. Computer monitoring has evolved to the point that some computer systems are designed to record and
evaluate information as detailed as the number of keystrokes made per minute and the number of breaks a computer user takes; other systems allow managers to access an employee's computer screen and watch work from a distance (Laabs, 1992).

Electronic performance monitoring has become one of the more controversial areas involved with new technology in the workplace. Critics charge that computerized employee monitoring is an invasion of privacy and disregards human rights, undermines trust, reduces autonomy, and emphasizes quantity to the exclusion of quality. Other critics suggest that electronic performance monitoring causes stress and is counterproductive because its use leads to declines in employee morale and productivity. Forester and Morrison (1990) described an organization in which the main computer records how long each clerk spends on each call and how much time separates one call from the next. Workers earn negative points for spending more than a predetermined target time handling a call, or taking more time than allowed for breaks.

Supporters of computer monitoring argue that it provides concrete, accurate performance measurement and a mechanism around which managers can motivate employees by providing incentives and effective rewards based on merit and effort. They also point out that what is being measured is factual, and that many workers favor such systems because they allow diligent workers to legitimately argue a case for better pay and benefits, with the case not relying on personal opinions or personalities (Forester & Morrison, 1990).

Data to support or refute the usefulness of the method are scarce. Laabs's (1992) interviews with management and employees subject to performance monitoring in a number of organizations (including Duke Power Company, AT&T, Toyota, Avis, and Charles Schwab) suggest that electronic monitoring is especially useful if its primary purpose is employee development. Managers in these organizations argued, for example, that a useful way to improve individuals' selling skills is to listen and critique their performance on sales calls. Some organizations listen in and provide immediate feedback in the form of noting strengths and weaknesses. Others (such as Duke Power Company), according to Laabs, rate performance on a nine-point scale across a number of dimensions such as courtesy, listening ability, empathy, job knowl-
edge, professional and positive attitude, communication skills, documentation and follow-up, marketing, and closing statement.

Murphy and Cleveland (1991) suggested that automatic monitoring may be regarded negatively by employees, but little is known about the conditions under which it will cause major problems (for example, employee stress or lowered morale). Turnage (1990) proposed that if it is used for evaluation, monitoring may cause distress and affect individual motives, goals, and performance; if it is used to give workers accurate, meaningful, nonevaluative feedback, however, it could serve to increase goal setting, allow greater control over individual performance, and thus increase motivation.

What additional influences could automatic monitoring of employees' behavior have on that behavior and performance appraisal? Grant, Higgins, and Irving (1988) examined the behavior and attitudes of monitored and unmonitored claims processors in a large insurance company. They found that monitored employees rated production quantity as the most important factor in their jobs, while unmonitored employees identified customer service and teamwork as their most important job factors. The authors suggested that monitored workers reasoned that unmonitored job factors did not count. A similarly negative outcome could be that supervisors will feel compelled to bring their evaluations in line with the objective data, even if they know that this depiction tells only part of the story regarding performance.

Grant and colleagues (1988) suggested that systems not monitoring all aspects of a job should be supplemented by other means of appraisal. They also noted that failure to provide frequent feedback on unmonitored performance dimensions will convince the employee that those dimensions are not important. These researchers also found that monitored employees were often aware that the performance data being collected through monitoring did not measure all parts of their job and perceived this as unfair.

As mentioned earlier, one of the consequences of new office communications technology is the ability to work at locations away from the traditional office. Computer monitoring may be especially appropriate as a means for gathering performance information about individuals who work in relative isolation or who spend much of their work week away from the traditional office. In addition, it may help
to keep individuals "honest" if they are asked to complete self-ratings. Clearly, much additional research is needed on this method before we can reach any definitive conclusions about its usefulness.

**Work Samples**

Historically, work samples or performance tests have been used more commonly as predictors than as criterion measures (for example, Asher & Sciarino, 1974). However, in one sense they represent a compelling criterion measurement method. What could be a fairer measure of job performance than to ask incumbents to complete several of the most important job tasks and then evaluate their performance on them? Performance tests appear to address directly the incumbent's ability to do the job. Performance or work sample testing has been referred to as a "high-fidelity" criterion measurement method, and other methods, such as job knowledge tests and performance ratings, as "surrogates" to performance tests (Widfor & Green, 1991). In fact, researchers (for example, Ghiselli & Brown, 1948; Guion, 1979; Robertson & Kundal, 1982; Siegel, 1986) have advocated work sample tests because they are direct, relevant measures of job proficiency that extract samples of behavior under realistic job conditions.

The most frequently used work sample methodology is hands-on testing. Individuals are asked to demonstrate proficiency on a set of tasks by performing the steps required for successful completion. Hands-on tests have been used as criteria for evaluating training programs (for example, Goldstein, 1974), validating selection devices (for example, Siegel & Jensen, 1955), identifying individual skill deficiencies (for example, Goldstein, 1980), and establishing worker job/task certification (for example, Guion, 1979).

The major advantage of using work sample methods is that they permit the employee's skills and knowledge to be compared to known standards under controlled and uniform conditions. Their main drawback is that the employee's performance during testing may not accurately reflect that person's daily performance on the job. In addition, work sample measures rarely cover all aspects of the criterion space.

Recent work by Hedge and Teachout (1992) on a measure of individual technical job competence may provide further support for the use of work sample methodology. Their approach, known as Walk-Through Performance Testing, is a task-level job performance measurement system that combines hands-on task performance and interview procedures. The hands-on component resembles a traditional hands-on work sample test designed to measure proficiency on a set of critical tasks. The interview component of the test requires incumbents to describe in detail how they would perform a job-related task. The administrator evaluates their show and tell descriptions for proficiency-based strengths and weaknesses related to the performance of that task. Just as with the hands-on component, the test administrator records correct or incorrect completion of each step of the task. One potential benefit of this interview testing methodology is that it combines the rigors of work sample testing with measurement efficiency (that is, time/cost savings). In addition, it affords the opportunity to assess proficiencies on tasks that cannot be measured feasibly in a hands-on mode.

In the future, as technologies continue to evolve rapidly, technical skills may become obsolete more quickly, and the need will increase for frequent evaluation of the employee's performance. However, because of their own skill obsolescence, supervisors may find it difficult to evaluate their subordinates' technical proficiency accurately. In such situations, a work sample criterion measure and "certified evaluators" may be an especially useful supplement to supervisory ratings for administrative or developmental purposes.

**Productive Capacity**

In 1989 the Air Force began to examine time-based measures of job performance (Carpenter, Monaco, O'Mara, & Teachout, 1989). The notion here was that a quality-based metric (that is, the usual 1–5, 1–7, 1–9 rating scale) was not as suited to addressing certain manpower issues as was a quantity-based metric. Consider, for example, the problem of estimating repair time for four F-111s with certain specified problems. How long will it take five average airman mechanics to do the work? Or conversely, given that we must get the four F-111s in the air in sixteen hours to meet mission requirements, how many average-performance-level airman mechanics will be required, how many high-performance-level mechanics, and so on?
The variables necessary for effective team performance (for example, Hackman, 1983, as described in Salas et al., 1992; Morgan, Glickman, Woodard, Bliwes, & Salas, 1986; Tannenbaum et al., 1992). For instance, the integrative model of Tannenbaum and others (1992) views team effectiveness as a product of multiple factors (examples are in parentheses): the organizational and situational context (organizational climate; reward systems), task characteristics (task complexity), work structure (team norms), individual characteristics (ability; motivation), team characteristics (team cohesiveness; member homogeneity), and team processes (coordination; communication). Outputs (and examples) are team performance (quantity; quality; errors), team changes (new norms; new communication patterns), and individual changes (attitudes; motivation). These models have much in common, and for our purposes serve to demonstrate the additional levels of complexity associated with team performance appraisal.

In some situations, team performance can be measured as an average of individual outputs, but in other situations such a strategy can be misleading. For example, Salas and others (1992) noted that individual skills are necessary for team success but are not sufficient for good team performance. In fact, Steiner (1972) defined "process loss" as occurring whenever team member efforts are wasted or duplicated in the course of meeting team coordination and communication requirements. Thus Steiner described team performance as a function of the ability of team members to perform their individual tasks, coordinate their work flow, and communicate effectively with one another.

The team evolution and maturation model developed by Morgan and others (1986) aptly describes this process as consisting of two basic tracks that must be successfully integrated for satisfactory team performance to be reached. One track deals with task-oriented skills that members must understand and acquire for task performance; the other track reflects the behaviors and attitudes that team members must develop before they can function effectively as a team. Specific teamwork behaviors include coordination, adaptation to varying situational demands, effective communication, compensatory behaviors, mutual performance monitoring, and giving/receiving feedback.

Team Appraisal

Interest in organizing work around teams has increased in the last few years. Generally, a team is defined as an interdependent collection of individuals, assigned specific roles and functions, who work toward a common goal or objective (see Salas, Dickinson, Converse, & Tannenbaum, 1992; Sundstrom, De Meuse, & Futrell, 1990). Various researchers have postulated models that describe
Individual performance within the team obviously can be appraised by any of the methods previously described. For example, supervisors could evaluate each of the team members under them. What is more interesting, and what may become more common in the future, is to evaluate the team’s performance. With the emergence of work teams, managers’ objectives increasingly may be tied to team goals. Consequently, it may be important to assess work group performance as well as individual performance. Overall team performance might be assessed, or multiple dimensions could be rated (for example, Morgan’s task-oriented and teamwork dimensions).

An innovative method developed by Pritchard, Jones, Roth, Stuebing, and Ekeberg (1988) provides a measure of team or unit overall performance by merging specific indicators into an index of percentage. Briefly, they had organization members identify multiple outcomes or products from the organization. Then, in an especially appealing feature of the method, they scaled different levels of the outcomes according to the degree of effectiveness represented. Consider, for example, a repair facility unit, one of the organizations for which Pritchard and colleagues developed a productivity measurement system. Organization members, with the help of a consultant, first identified seven indicators that they believed reflected reasonable indexes of the unit’s productivity. For example, one of the indicators was the number of units per week returned for re-repair. The unit members then constructed for each indicator what the researchers called a contingency, as illustrated in Figure 12.1.

In the example in Figure 12.1, unit members believed that five returns per week are expected, essentially an average outcome, and this was assigned an effectiveness score of 0. They thought the best possible return rate was two per week, and assigned a +80 score to that outcome; they judged the worst realistically possible to be ten per week and assigned a -70 effectiveness score. The group then scored the remaining outcomes, resulting in the contingency appearing in Figure 12.1. As is evident from the example, contingencies may be nonlinear. Here two returns per week is not much more effective than three per week. The importance of the outcome to organizational effectiveness is represented by the slope of the curve, with a steep slope indicating a relatively important outcome. Using the contingencies, organizations can score individual units on their outcomes and pool these scores to provide a single effectiveness value. The main advantage of this method is that multiple indicators of productivity can be pooled in a meaningful and psychologically sound way.

Pritchard and colleagues (1988) have used the method primarily in productivity improvement programs, but it may also be employed in group performance appraisal. Different contingencies can be generated for different performance dimensions, and overall effectiveness scores computed for individual units. This method ties in nicely with the notion of rewarding performance at the team rather than the individual level. For example, assigning pay or bonuses based on appraised group performance should foster teamwork and encourage cooperation and support among team members. With increasing emphasis on teams in organizations,
group performance appraisal, using Pritchard’s approach or similar strategies, should become more common.

Other Factors to Consider for the Future of Performance Measurement

In the twenty-first century, given the demands that will be placed on individual and team performance in the workplace, other criterion concepts and related areas of research should be of increasing concern. We discuss in this section the notions of maximal and typical criteria, technical proficiency versus contextual criterion domains, attitudes about performance appraisal, and selecting criterion measures.

Maximal Versus Typical Performance

Many years ago Cronbach (1960) made the useful distinction between maximal and typical performance. He referred to maximal, “can-do” performance as ability-related, and typical, “will-do” performance as driven more by motivational factors than by ability. Current and future performance appraisal practice should take into account this distinction.

Campbell’s (1990) model of the determinants of job performance clarifies the distinction. The model implies that performance is a function of declarative or factual knowledge, procedural knowledge (that is, knowing how to do a task), and motivation. Maximal performance is involved in the first two components, and the motivation component is important for typical performance. Maximal performance measures, such as work samples and hands-on performance tests, usually constrain those tested to try hard for the short duration of the test, and thus motivation is essentially held constant. Typical performance, on the other hand, depends substantially on motivation. Will-do performance-over-time requires job knowledge certainly, but also requires sustained, motivated effort in a setting where motivation is not constrained and can clearly vary across job incumbents.

How does ability fit in here? Performance models offered and confirmed by Hunter (1983), Schmidt, Hunter, and Outerbridge (1986), and Borman, White, Pulakos, and Oppler (1991) demonstrate a clear path from ability (that is, general cognitive ability) to job knowledge to technical proficiency, where the proficiency variable is a maximum performance measure. Interestingly, when Borman and associates included supervisory performance ratings with selected personality measures and behavioral variables reflecting mostly typical performance, a model emerged with achievement orientation, dependability, and technical proficiency all influencing the ratings (see Figure 12.2).

One way to interpret these results, within the framework of maximal/typical performance and their antecedents, is as follows. The ratings are likely measures of both maximal and typical performance. They are meant to tap typical performance, but raters may consider can-do performance when rating such dimensions as Technical Knowledge and Technical Skill. Accordingly, maximal performance (technical proficiency) appears to be a function of ability and job knowledge, but not so much a function of personality. Typical

---

Figure 12.2. Antecedents of Maximal and Typical Performance.

![Diagram showing antecedents of maximal and typical performance]

Source: Adapted from Borman, White, Pulakos, & Oppler, 1991.
performance, as captured in the ratings, has as antecedents both the ability/job knowledge/technical proficiency sequence of variables and personality through the typical performance behavioral variables. It may be that the technical proficiency-ratings path would not be as large if the ratings reflected only typical performance, although this is clearly speculation. Nonetheless, what this performance model suggests is that maximal and typical performance can be distinguished because they have somewhat different antecedents. They are both very important for organizational functioning and may not be highly correlated.

A study by Sackett, Zedeck, and Fogli (1988) provides a more direct picture of the maximal-typical performance linkage. Working with grocery clerks, these researchers designed a maximal performance work sample test that consisted of ringing up items in a standardized shopping cart. Measures of speed (total time to complete the cart) and accuracy (number of voids and incorrect entries) were derived from this test. For the typical performance measure, a computer-monitored system kept track of each grocery clerk’s performance over a thirty-day period. Speed (number of items rung up per minute) and accuracy (number of voids per day) were measured. The most striking finding here is the generally low correlations between these two components of performance. A follow-up study (Du Bois, Sackett, Zedeck, & Fogli, 1993) demonstrated a somewhat different pattern of predictors of maximal and typical performance across two samples, but once again, ability-related predictors were linked more substantially to maximal performance than to typical performance.

The point we wish to make from this discussion of typical versus maximal performance is that the researcher or practitioner needs to consider what kind(s) of performance should be evaluated in the particular application, and understand that a somewhat different picture of performance and predictors of performance may emerge from the two types of criteria. The latter point is important for the future because as personality and related tests become more acceptable as selection instruments, they may be increasingly used when typical performance is the target criterion domain. That is, appreciation of the maximal-typical performance distinction should make it even more likely that personality predictors are seen as appropriate for personnel selection because of their link with typical, will-do performance.

Technical Proficiency Versus Contextual Criterion Domains

Another construct distinction is between task performance and contextual performance. Borman and Motowidlo (1993a) discussed this distinction and its implications for personnel selection. Task performance relates to the proficiency with which job incumbents perform the core technical activities that are important for their jobs. The authors defined contextual performance as extra-task proficiency that contributes more to the organizational, social, and psychological environment to help accomplish organizational goals. Integrating elements of organizational citizenship (for example, Organ, 1988), prosocial organizational behavior (for example, Brief & Motowidlo, 1986), and a model of soldier effectiveness (Borman, Motowidlo, & Hanser, 1983), Borman and Motowidlo (1992) identified the following contextual dimensions:

- Persisting with enthusiasm and extra effort as necessary to complete one’s own task activities successfully
- Volunteering to carry out task activities that are not formally part of one’s own job
- Helping and cooperating with others
- Following organizational rules and procedures
- Endorsing, supporting, and defending organizational objectives

There are conceptual and some limited empirical arguments to suggest that contextual performance is related to organizational effectiveness. That is, organizations with members effective in the contextual performance domain will tend to be more successful than organizations whose incumbents are ineffective in this domain (Borman & Motowidlo, 1993a). A recent study of U.S. Air Force enlisted personnel by Motowidlo and Van Scotter (described in Borman & Motowidlo, 1993b) found that the correlation for task and contextual performance was only .17, indicating that contextual performance cannot be “covered” by measuring technical proficiency alone.

Our view is that contextual performance will be seen as more important in the future. As the global economy demands that organizations become even more competitive, more will be expected of employees. As work teams become more prevalent, cooperation and working smoothly with others will be increasingly important.
Finally, as customer service continues to be a major emphasis in organizations, going “above and beyond the call” for customers will be increasingly expected of organization members.

Attitudes About Performance Appraisal

Workers’ desires for autonomy and self-development are increasing, and today’s workers have a growing sense of entitlement to be involved in the decisions pertaining to their work (Offermann & Gowing, 1990). Consequently, worker attitudes toward performance appraisal may play an increasingly important role in performance appraisal in the future. Murphy and Cleveland (1991) noted that the dominance of psychometric and accuracy criteria have diverted researchers’ attention away from three classes of criteria that might be critical in determining the success of an appraisal system: reactions, practicality, and decision processes. They argued that reaction criteria (such as perceptions of fairness and accuracy of appraisal systems) probably place a ceiling on the effectiveness of the system; that is, acceptance of the system by raters and those rated may be necessary but not sufficient for the system to be effective. Dickinson (1993) suggested that if negative attitudes about performance appraisal prevail among organizational members, its use may hinder rather than help achieve outcomes.

In spite of the commonsense logic that acceptance of a personnel procedure is crucial to its effective use, the notion that attitudes toward performance ratings could affect their validity was not noted until 1967, when Lawler proposed a model of the factors that affect the construct validity of ratings. Central to the model was the belief that attitudes toward the equity and acceptability of a rating system are a function of organizational and individual characteristics as well as the rating format.

Landy, Barnes, and Murphy (1978) were among the first researchers to relate attitudinal factors empirically to job performance measurement. They identified four significant predictors of perceived fairness and accuracy of performance appraisals: (1) frequency of appraisal, (2) plans developed with the supervisor for eliminating weaknesses, (3) supervisor’s knowledge of the job duties of the person rated, and (4) supervisor’s knowledge of the level of performance of the person rated. In a follow-up study with the same population, researchers found that the level of the performance rating did not affect these relationships (Landy, Barnes-Farrell, & Cleveland, 1980). Dipboye and de Pontbriand (1981) further distinguished employees’ opinions as to whether they related to the performance appraisal system or to the appraisal itself. They found that four factors related to the two dependent variables: (1) favorability of the appraisal, (2) opportunity for employees to state their own perspective in the appraisal interview, (3) job relevance of appraisal factors, and (4) discussion of plans and objectives with the supervisor.

A series of studies by Kavanagh and colleagues (Kavanagh & Hedge, 1983; Kavanagh, Hedge, Ree, Earles, & De Biasi, 1985) found that several attitudes toward the appraisal system were significant predictors of appraisal acceptability across studies. These included attitudes about whether (1) the appraisal system facilitates fair and accurate appraisals, (2) the appraisal system allows raters to distinguish between workers’ proficiencies, (3) the appraisal system provides clear performance standards, (4) the individuals rated receive satisfactory feedback, and (5) they receive a satisfactory performance evaluation.

A related body of literature, which has received renewed attention of late with its translation into performance appraisal terms, concerns organizational justice (Thibaut & Walker, 1975). This research is directed at identifying the features of organizational procedures that affect perceptions of fairness, work attitudes, and behavior. The literature suggests that there are two dimensions of perceived justice: distributive justice and procedural justice. Interpreted in performance appraisal terms, distributive justice focuses on the fairness of the evaluations received relative to the work performed. Procedural justice focuses on the fairness of the evaluation procedures used to determine the ratings (Greenberg, 1986).

Giles and Mosholder (1990) identified four important organizational actions related to satisfaction with appraisals: removing excessive complexity from appraisal systems, properly introducing them into organizations, ensuring that supervisors conduct appraisal sessions, and linking salary adjustments with appraisal results. They suggested that if such actions are neglected, it is likely that the impact of the appraisal system will be substantially reduced.

A number of years ago Jacobs, Kafry, and Zedeck (1980) noted their own disappointing experiences with organizations that abandoned recently developed appraisal systems. They suggested that
many organizations revert to evaluation systems in use prior to intervention because of organization policy and the excessive personnel time and energy requirements associated with, for example, behaviorally anchored rating scales. These frustrations, in a very applied way, speak to the issue of acceptability, and suggest its importance as a criterion when evaluating the effectiveness of an appraisal system. After all, if a psychometrically sound system is unacceptable to its users, it might never be used, or it might be used improperly. As Banks and Murphy (1985), and Longenecker, Sims, and Gioia (1987) have noted, raters must not only be capable, but they must also be willing to provide accurate ratings.

As the future work environment pushes us toward use of new and different appraisal techniques, it would be wise not to lose sight of how these approaches may be perceived by all parties involved in the appraisal process. Acceptance of a particular method or methods may be critical for successful implementation and use.

Identifying Additional Criteria for Criterion Measures

Over the years, researchers have identified and/or developed many examples of “criterion for criteria,” or standards on which to assess the quality of criterion measures (Weitz, 1961). Beilords (1954) suggested that criterion measures be reliable, realistic, representative, related to other criteria, acceptable to the job analyst, acceptable to management, consistent from one situation to another, and predictable. Blum and Naylor (1968) proposed that criterion measures should also be inexpensive, understandable, measurable, relevant, uncontaminated and bias-free, and discriminating. Bernardin and Beatty (1984) compiled a large list of variables and clustered them into three primary categories of criteria: quantitative (for example, reliability, validity, discriminability); utilization (for example, feedback, merit pay, adverse impact); and qualitative (for example, amount of documentation, user acceptability, maintenance costs).

Although researchers have periodically called attention to the importance of using multiple criteria to judge criterion measures, they have seldom provided operational definitions of these variables. If some form of empirical evaluation is included, it is typically dominated by considerations of reliability and validity. Several researchers have attempted to apply a more systematic process to the use of multiple criteria to judge performance measures.

McAfee and Green (1977) applied sixteen criteria to aid the selection of a performance appraisal method for use in a large midwestern hospital. They rated ten different appraisal methods on criteria such as usefulness for counseling and employee development, expense to develop, reliability, and freedom from psychometric errors. They used a weighted sum to identify the best method for the job and organization under consideration. Drawing on this work, Kavanagh (1982) proposed nineteen criteria, each operationally defined, against which to judge the value of performance appraisal systems. Included were such things as: psychometric quality, developmental costs, user acceptance, periodic review/feedback, meeting guidelines of the Equal Employment Opportunity Commission (EEOC), and susceptibility to inflation of ratings.

As dramatic changes continue to occur in the structure and function of the workplace, changes may also be required in how we approach the measurement of job performance. In this chapter we have suggested that performance appraisal in the workplace of the future may, by necessity, become much more varied and multifaceted. As such, reliance on a more formalized process of applying the type of criteria described by McAfee and Green (1977) and Kavanagh (1982) for selecting performance measures should improve the chances that adequate measures are chosen and that accurate, useful performance information is obtained.

Conclusion

Our focus in this chapter has been on the need for, and use of, performance appraisal strategies for enhancing the effectiveness of organizations. We believe that efforts devoted to evaluating performance are critical for workforce productivity. The measurement of employees has played and is likely to continue to play a major role in organizations, unless future jobs get structured to (1) eliminate performance variability or (2) severely reduce the control over performance by the worker.

Consequently, as we move toward the twenty-first century, rapidly changing job requirements and job environments demand that our performance appraisal systems reflect these changes. The
performance appraisal system can help connect employees to the organization during these changing times; it can help them to identify how changes will affect their jobs, how performance expectations/goals will change, and how they will need to adapt to changing job requirements.

Regarding research, we must be open to investigating different measurement methods such as electronic performance monitoring, team effectiveness indexes, and subordinate and customer ratings. Research on objective performance measures and supervisor and peer ratings will remain important, but some of our attention should be focused on emerging methods as well.

Many organizations wisely employ a variety of strategies to increase individuals’ productivity and organizational effectiveness. From the perspective of this chapter, we conclude by noting that only through carefully developed and implemented performance appraisal programs can we hope to utilize human resources efficiently in changing organizations.

References


