PERFORMANCE MANAGEMENT AND ASSESSMENT: METHODS FOR IMPROVED RATER ACCURACY AND EMPLOYEE GOAL SETTING

Manuel London, Edward M. Mone, and John C. Scott

This article examines the gaps between research and practice in the areas of rater accuracy and goal setting. Prior research has shown that human resource managers may incorrectly believe that training raters to recognize errors will increase rater accuracy and that employee participation in goal setting is more effective than assigning goals. Theory-based research suggests ways to help raters recognize expected performance and enable employees to self-regulate their pursuit of goals. We describe applications of these findings to performance management programs and suggest methods for evaluating their effectiveness once implemented. © 2004 Wiley Periodicals, Inc.

Introduction

Performance management systems in organizations should rest on principles of good management supported by extensive research in organizational psychology and human resource management. Basic methods of performance management encompass dimensions of performance, goals for performance and career development, training and development programs, supervisory performance appraisal, multisource performance ratings, feedback, coaching, and rewards. Unfortunately, human resource practitioners and managers may misunderstand or not be aware of research on these central elements of performance management systems. In fact, they may have erroneous views that do not stem from research at all. Managers have an excuse since they were not trained in human resources. Practitioners don't.

This article focuses on two key elements of performance management systems: rater accuracy and goal setting. We begin by highlighting gaps between research and practice in these areas pointed out by Rynes, Colbert, and Brown (2002). These gaps suggest problems with how organizations manage and assess employee performance. We then review theory and research on rater accuracy and goal setting. In the final section, we suggest how to design and assess these key elements of performance management programs, and related elements such as feedback and coaching, to ensure that they meet current and emerging organizational needs and re-

Correspondence to: Manuel London, PhD, State University of New York at Stony Brook, Harriman Hall, Stony Brook, NY 11794-3775, 631-632-6833, e-mail: mlondon@notes.cc.sunysb.edu

flect new knowledge in organizational psychology and human resource management.

What Practitioners Don't Know

Regarding rater accuracy in performance appraisal and the value of rater training, 75% of Rynes et al.'s (2002) sample *incorrectly* believed that most errors in performance appraisal can be eliminated by providing training that describes the kinds of errors managers tend to make and suggesting ways to avoid them. Perhaps this kind of training is viewed as a panacea to cure all ills, but how can practitioners be fully aware of the kind of errors raters make and the extent to which these errors affect rater accuracy? Regarding employee participation and goal setting, 82% of their sample *incorrectly* believed that, on average, encouraging employees to participate in decision making is generally more effective for improving organizational performance than establishing performance goals. The presumed value of participation appears to be ingrained in the zeitgeist of what HR professionals think is good management practice. Isn't goal setting, participation in decision making, and objective feedback what management by objectives (MBO; Drucker, 1954; Rodgers & Hunter, 1991) is all about, and isn't this a tried and true management technique?

Oddly, research on the topics of rater error and goal setting are in the mainstream of organizational psychology. This research, however, has not infiltrated the minds and practice of human resource managers, as evidenced by the recent increased attention in the HR practitioner and management literatures regarding HR systems often being deficient and even demotivating. There are several probable reasons for this:

- The misuse of methods, poor program development, and lack of program evaluation, or simply that researchers may not be clearly communicating the practical aspects of their results.
- HR practitioners may not recognize the context—what's happening in their organizations that can influence the application of this research.

They may not understand and apply evaluation methods, including pilot testing, ongoing tracking of results, and measurement of independent and objective outcomes.

- HR practitioners might not be choosing the right methods for the right purposes. For instance, organizations may turn to ratings from subordinates (upward ratings) and multisource (360-degree) feedback (ratings from subordinates, peers, and customers) to supplement ratings from the immediate supervisor. However, these methods of collecting ratings from sources other than the supervisor were designed for development, and their use for administration reduces their reliability and validity (Greguras, Robie, Schleicher, & Goff, 2003). The larger problem may be trying to combine the two uses of multisource ratings—feedback and decision making—when the two functions run at cross-purposes.
- HR practitioners might not be choosing the appropriate criteria to evaluate performance management programs. Organizations may expect that these programs will affect bottom-line outcomes at the organization or company level, outcomes that are influenced by many factors, while the programs were designed to affect individual learning and behaviors that have only indirect and long-term effects on bottom-line outcomes.
- New appraisal and goal-setting methods may be implemented quickly with little attention to interpersonal dynamics and psychometric testing and refinements to ensure reliability and validity. These methods may not be evaluated to determine whether they are operating as intended in comparison with the methods they replaced or control groups.

Research Findings on Rater Accuracy and Participation in Goal Setting

The research on rater accuracy and participation in goal setting is not new, but it is

worthwhile to review in light of Rynes et al.'s (2002) findings to better understand why these areas are not well known or understood. We do not attempt to review the large body of empirical research on performance appraisal or goal setting. Instead, we focus on key empirical/theory-based findings and how they generalize to what can be put into action. Table I summarizes these findings and their implications.

Rater Training in Performance Appraisal

Hedge and Kavanagh (1988) noted that early approaches to rater training were concerned with eliminating rating errors, such as leniency, halo (letting a positive evaluation on one performance dimension influence ratings on other performance dimensions), and range restriction (using only one portion of the scale—for instance, just the middle categories). Rater error training (RET) cautioned raters to avoid these errors and was effective in doing so. However, RET may teach raters to use inappropriate response sets, such as being sure to use the full range of the response scale, even when this isn't appropriate (Bernardin & Buckley, 1981). On the other hand, Hakel (1980, cited in Hedge & Kavanagh, 1988) noted that rating errors might not be errors at all. Hakel suggested that rating errors should be called rating effects, since a rating effect, such as leniency, may indeed be an accurate reflection of an

---

TABLE I **Summary of Findings and Implications for Practice**

Source	Finding	Application
Rater Training in Performance Appraisal		
Hedge & Kavanagh (1988) Bernardin & Buckley (1981) Hakel (1980)	Rater error training may teach raters to use inappropriate response sets.	Rater errors should be called rating effects, which may be accurate reflections of performance; reducing rating effects may decrease rating accuracy.
Bernardin & Buckley (1981) Woehr (1994)	Identify norms for effective performance behaviors as standard frames of reference.	Frame-of-reference training provides examples of good, average, and poor performance for each behavioral dimension. It helps raters identify and classify performance correctly.
Sulsky & Balzer (1988) Murphy & Balzer (1989)	The average correlation between rating effects and direct measures of accuracy is low.	Do not use rating effects as indirect measures of accuracy.
Stamoulis & Hauenstein (1993)	Frame-of-reference training helps distinguish between dimensions; rater error training increases accuracy of rating favorability; both forms of training improve discrimination among ratees.	Frame-of-reference and rater error training can be beneficial for different purposes. (See Table II.)
DeNisi & Peters (1996)	Structured diary keeping and recall increased ratees' positive reactions, recall, and accuracy of rating level and discrimination between and within ratees.	Training methods should help raters organize information into clear performance categories and store information by person and category.
DeNisi (1996)	Determine performance theories that guide raters' evaluations and best fit their organizations, and train raters to apply these theories to performance judgments.	Help raters understand expected performance and how to spot and record behaviors that reflect that performance.

(continued)

TABLE I **Summary of Findings and Implications for Practice**

Source	Finding	Application
Goal Setting and Participation		
Locke & Latham (2002)	Participation in goal setting does not affect performance when goal difficulty is held constant.	Be sure goals are challenging, whether individuals participate in setting their own goals or the goals are assigned.
Latham & Saari (1979) Latham & Marshall (1982) Latham, Steele, & Saari (1982)	Setting specific, challenging goals results in higher performance than urging people to do their best.	Be sure goals are specific.
Latham, Erez, & Locke (1988)	Goals that are assigned without explanation lead to lower performance than goals that are set participatively.	Explain the reason for goals that are assigned.
Locke, Alavi, & Wagner (1997)	Participation in goal setting has cognitive value by increasing information exchange about how to achieve goals and thereby increasing self-confidence.	Involve employees in discussions about goal difficulty and ways of achieving goals.
Latham & Yukl (1976) Latham, Mitchell, & Dosset (1978)	Participation in goal setting results in setting more difficult goals; this, in turn, increases performance.	Involve employees in setting goals because they will set more difficult goals for themselves, and goal difficulty will prompt more effort.
Locke & Latham (1990) Klein, Wesson, Hollenbeck, & Alge (1999)	Commitment to goals is related to performance; incentives, competition, and leader's power increase commitment. Goal commitment enhances the positive effect of goal difficulty on performance.	Focus on ways to increase goal commitment—e.g., establish rewards for goal accomplishment.
Locke (1991) Locke & Latham (2002)	Goal motivation increases with participation in goal setting, commitment to the goals, and belief that the goals can be accomplished.	Involve employees in setting goals to increase their commitment and self-efficacy.
Latham (2002) Locke & Latham (2002)	Feedback helps goal achievement.	Provide feedback to help employees calibrate their progress toward a goal.
Ilgen, Fisher, & Taylor (1979) London (2003)	Feedback focuses attention on goals, helps discover errors, maintains goal direction, and provides information on capabilities and effort needed for goal achievement.	Provide feedback to enhance goal achievement.
Locke & Latham (2002) Bandura (1997)	Training programs to enhance self-regulation teach employees to set specific challenging goals and monitor how their environment affects performance.	In designing performance management programs, include training in self-regulation to help employees overcome barriers.
Oettingen, Pak, & Schnetter (2001)	Self-regulation reduces decision time, increases planning, controls attention, and reduces impulses that are not task-related.	Support and reinforce self-regulation to maintain goal direction, attention to effort needed to accomplish goals, and recognition of goal outcomes.

employee's performance. So although training to reduce the rating effect may reduce errors, it may also result in inaccurate ratings.

One way to help raters make correct behavioral observations and increase their rating accuracy is called frame-of-reference (FOR) training (e.g., Bernardin & Buckley, 1981). First, norms for effective performance behaviors are identified, and these norms become the standard frame of refer-

ence. FOR training, which provides raters examples of good, average, and poor performance for each behavioral dimension, helps raters identify and classify observed performance correctly. As a result of this training, raters develop shared performance schema for processing information (Woehr, 1994), and the schema help increase the recall of behavioral performance information and produce more accurate ratings. A second way to help raters is through training that focuses on the judgments or decisions raters make when evaluating the behaviors they observe. In this approach, raters listen to a videotaped lecture and learn about intuitive and formal decision-making strategies, as well as the costs and benefits of formal versus intuitive strategies. They also learn about inferential errors, such as inappropriate causal inference or giving more weight to a behavior than it deserves.

Hedge and Kavanagh (1988) compared rater error (response set) training, observation training, and decision-making training. They found that traditional rater error training reduced rating effects, but it also had a negative effect on rating accuracy (the correlation between the rater's ratings on a dimension for all employees rated and the mean ratings provided by expert judges). Observation and decision-making training increased rating accuracy but did not reduce the classic rating effects.

While rating effects are indirect measures of rater accuracy, direct measures of accuracy include the overall level of ratings (called elevation); discriminating among employees (differential elevation); discriminating among dimensions of job performance (stereotype accuracy); and discriminating among ratees within dimensions (differential accuracy) (derived from Cronbach, 1955; see Sulsky & Balzer, 1988, for a review of performance appraisal studies using these accuracy measures). Murphy and Balzer's (1989) meta-analysis found that the average correlation between rating effects (an indirect measure) and Cronbach's direct measures was extremely low (.05). They recommended not using rating effects as indirect measures of accuracy, since there is no way to tell whether observed rating effects are too large, too small, or correct,

since the measures in and of themselves are not compared to true ratings or scores.

Different aspects of accuracy may be useful for different purposes. Elevation and differential elevation are the most important measures of accuracy for making inter-individual decisions because they drive and validate personnel decisions. Elevation and differential elevation may not be the most valuable measures of accuracy when the goal of the measurement process is accurate assessment of individual strengths and weaknesses. Discriminating among ratees within dimensions (differential accuracy) is important for the accuracy of personnel placement decisions (Murphy & Balzer, 1989).

Later research found that raters who received FOR training were the most accurate in discriminating among behavioral dimensions of job performance (stereotype accuracy) and discriminating among ratees within dimensions (differential accuracy), while those who received RET were most accurate on elevation (level of ratings), and raters who received either RET or FOR training improved their ability to discriminate among ratees (Stamoulis & Hauenstein, 1993). Therefore, rater error training would be more useful when the primary goal of appraisal is to distinguish accurately among workers, while frame-of-reference training would be more useful when the primary goal of the appraisal system is to provide accurate developmental feedback to workers.

DeNisi and Peters (1996) observed that studies dealing with cognitive processes in performance appraisals, such as those cited above, were conducted in laboratory settings, often with college student subjects. They conducted field experiments in two organizations to evaluate the effects of two interventions aimed at helping raters to better organize performance information in memory. The two interventions were structured diary keeping and structured recall (structured recall asks raters to record, from memory, all critical incidents of performance they observed). They found that the two interventions increased raters' positive reactions to the appraisal process, enhanced their ability to recall performance information, and resulted in ratings that were less elevated and more dis-

Later research found that raters who received FOR training were the most accurate in discriminating among behavioral dimensions of job performance...

criminating between and within ratees. Structured diary keeping and structured recall work in the same way as FOR training to help raters organize information. While FOR training provides raters with performance categories for organizing information before they actually acquire it, diary methods and structured recall help raters store information by person or task (that is, performance dimension) or organize information after it has been encountered and encoded but before it is actually used in making ratings.

DeNisi (1996) concluded that rater error training does not increase rater accuracy and that a more effective approach may be to study performance theories that guide raters' evaluation processes, determine the theories that best fit their organizations, and train raters to comprehend and apply those theories. DeNisi also noted that the quality of ratings can be improved if raters know the performance expectations and job behaviors desired by the organization and if they are provided formats to gather and interpret performance information and make judgments about it. This may be a subtle distinction. That is, rater training can be valuable, but don't educate raters about the errors they make. Rather, help raters understand the nature of expected performance and how to spot and record behaviors that reflect that performance. By the way, we do not mean to imply that rater training is the solution to all accuracy problems. Rating format and rating conditions (e.g., the opportunity to observe relevant behaviors, incentives/disincentives to providing accurate ratings) may affect the ability and willingness of raters to provide accurate ratings. The tools raters use and the social forces that encourage or, more often, discourage accurate ratings are also important.

Table II highlights the purposes, content, weaknesses, and uses of different rater training methods. The table emphasizes that different training methods affect different conceptualizations of accuracy and, therefore, have different uses.

Goal Setting and Participation

The idea that goal setting is likely to be more effective when people participate in setting

...rater training can be valuable, but don't educate raters about the errors they make. Rather, help raters understand the nature of expected performance and how to spot and record behaviors that reflect that performance.

goals than when goals are assigned to them stems from the notion that participation increases commitment to the goal. The presumption is that people who participate in setting goals are likely to be more motivated to achieve them than those who are given goals created by others. In a review of literature on this topic, Locke and Latham (2002) reported that when goal difficulty is held constant, the performance of those who participate in setting goals does not differ significantly from the performance of those who were assigned goals. In fact, both a participatively set goal and an assigned goal resulted in higher performance than merely urging people to do their best (cf. Latham & Marshall, 1982; Latham & Saari, 1979; Latham, Steele, & Saari, 1982). However, an assigned goal is only as motivating as one that is set participatively as long as the reason for the goal is given; if the goal is assigned without explanation, it leads to lower performance than a participatively set goal (Latham, Erez, & Locke, 1988). Locke, Alavi, and Wagner (1997) discovered that the benefit of participation in goal setting is cognitive, not motivational. Participation engenders information exchange, for instance, about strategies to achieve the goal and such discussions increase self-confidence in the ability to do so. However, participation in goal setting results in setting more difficult goals than those who are assigned goals (Latham & Yukl, 1976; Latham, Mitchell, & Dosset, 1978), and more difficult goals, not necessarily having participated in setting the goals, leads to higher performance.

Although participation doesn't have the expected effect on goal commitment, goal commitment can be increased in other ways. Employees' commitment to assigned goals can be increased by incentives, competition, and the leader's power (Locke & Latham, 1990). Commitment to assigned goals is important because it increases the relationship between goal difficulty and performance. A meta-analysis of 83 independent studies revealed that goal commitment has a strong, positive influence on performance (Klein, Wesson, Hollenbeck, & Alge, 1999). Furthermore, the meta-analysis revealed that goal commitment is higher when people find

| TABLE II | Rater Training—Purposes, Liabilities, and Strengths | | | |

Type of Rater Training	*Primary Purpose*	*Content of Training*	*Liability of Approach*	*Primary Value*
Rater Error (RET)	Conscious awareness of possible errors; discriminating among ratees (differential elevation)	Focus, e.g., on leniency, halo, range restriction	May use inappropriate response sets, may be a "rating effect," can reduce the rating effect but may also lower accuracy	Use to make decisions that require distinguishing accurately among employees—e.g., compensation decisions
Frame-of-Reference (FOR)	Develop shared performance schema; discriminating among dimensions of performance (stereotype accuracy) and discriminating among ratees within dimensions (differential accuracy)	Examples of normative (poor, average, good) behaviors for behavioral dimensions	Does not reduce rating effect	Use to make decisions that require comparing employees on different performance dimensions—e.g., job assignments and placements, and feed-back for development and goal setting
Decision-Making Strategies	How to gather, retrieve, and weight information about performance to make decisions	Formal and intuitive strategies	Focuses more on the application of performance information than its accuracy	Increases awareness of decision-making strategies; use to apply performance informa-tion to make decisions about employees
Structured Diary Keeping and Structured Recall	How to store performance information by person or trait and how to organize performance data after it is encoded	Establishing and using a structure for the ongoing recording of performance data and retrieving it when needed	Employees need to be disciplined to record observations and critical incidents when they happen and be able to recall details; focuses on content of information, not how to evaluate it	Use for feedback discussions about specific behaviors and performance outcomes; also use to gather information
Performance Theory	How to identify and apply the organization's and rater's theory of performance to increase accuracy	Communicating the performance expected and desired behaviors; provide formats for interpreting performance and making judgments **<zaq1>**	Focuses more on understanding differences between dimensions (stereotype accuracy) than distinguishing between employees (differential elevation) and understanding differences between employees on different dimensions (differential accuracy)	Use for feedback and goal-setting discussions that require differentiating elements of performance

the goal to be attractive and likely to lead to positive outcomes, and that goal commit-ment is not highly related to people's abilities or desires.

The "motivational hub" of goal setting consists of personal (self-set) goals, goal commitment, and self-efficacy (an individ-ual's belief that he or she can bring about positive outcomes) (Locke, 1991; Locke & Latham, 2002). These are ". . . the most im-mediate, conscious motivational determi-nants of action" (Locke & Latham, 2002, p.

709). Although assigning a challenging goal may raise self-efficacy, which, in turn, increases ownership of a goal and performance, self-set goals and self-efficacy will be more highly related when goals are self-set than when they are assigned <zaq2> (Locke & Latham, 2002).

Locke and Latham's (1990) goal-setting theory holds that specific and challenging goals lead to higher performance than no goals or setting general goals, such as "try your best." Also, as suggested above, Locke and Latham's (1990) theory holds that people who participate in setting their own goals are likely to set more difficult goals than others will set for them, and that goal difficulty leads to increased commitment to achieving the goals. Feedback and competition have a similar effect on performance. Therefore, the extent to which goals lead to performance depends on participation, commitment, and other elements of the performance management process and the situation, such as feedback (Latham, 2002).

A word about feedback may be useful here, because it plays a key role, along with goal setting, in the self-regulation of performance. Goals and feedback work together to affect goal accomplishment. Employees may provide their own feedback or receive it from others, such as supervisors or peers. Coaches are another source of feedback and support for participation in goal setting. Employees need feedback to help calibrate their progress toward a goal, as well as to suggest ways to adjust the level or direction of their efforts or to shift performance strategies. The combination of goals plus feedback is more effective than goals alone (Locke & Latham, 2002). However, feedback or praise in the form of public recognition or a monetary bonus do not affect performance unless they lead to setting, and committing to, specific and difficult goals (Latham, 2002). Overall, feedback focuses attention on performance goals that are important to the organization, helps discover errors, maintains goal direction, influences new goals, provides information on performance capabilities and how much more effort/energy is needed to achieve goals, and provides positive reinforcement for goal accomplishments

(Ilgen, Fischer, & Taylor, 1979; London, 2003).

Locke's goal-setting theory holds that self-regulation is a key to goal setting (Locke & Latham, 2002). Training programs to enhance self-regulation teach employees to set specific challenging goals, monitor ways in which their environment helped or hindered achievement of their goals, and use verbal self-guidance (Locke & Latham, 2002). Such training in self-regulation enables employees to cope effectively with personal and social barriers to goal accomplishment. This is consistent with Bandura's (1997) social cognitive theory in that the training increased self-efficacy, which enhances goal motivation. Encouraging verbal self-guidance and mental practice (symbolic guided rehearsal of a task before actually doing it) are ways to enhance an individual's self-efficacy.

Employees who operate in a self-regulatory mode in pursuit of a goal take less time to make decisions related to the goal, plan and initiate their actions, control their attention, and inhibit troubling impulses that might otherwise prevent them from sticking with a challenging task (Oettingen, Pak, & Schnetter, 2001). An intervention to help employees maintain their goal-directed behavior is mentally contrasting fantasies about a desired future with aspects of present reality that stand in the way of reaching the desired future (Oettingen et al., 2001). This expectancy-based approach directs what to do and how hard to try.

Table III summarizes different goal-setting approaches: participation, assigning goals, and urging to do one's best.

Research-Based Applications

In this section, we provide examples of approaches to rater training and goal-setting processes that derive from research.

Rater Training

Given the low costs associated with online training and Web-based resources, organizations are able to provide a breadth and depth of resources to support the performance

The "motivational hub" of goal setting consists of personal… goals, goal commitment, and self-efficacy…

TABLE III	Goal-Setting Approaches		
Approach	*Impact on Performance*	*Reason*	*Impact on Goal*
Participation	• Motivation increases • Performance improves	Information exchange results in cognitive benefits; self-confidence in ability is increased	More difficult goals are set
Assigning Goals	Performance improves, but only if reason for goal is given	Increase in performance is partially a result of participation	Goals set are not as difficult as when set with participation
Urging To Do Best	Not the highest possible level of performance	No participation or information exchange	Not as difficult

management process, often with the support of human resource managers who can provide face-to-face consultation when needed. For example, one high-tech firm created a robust performance management Web site on the company's intranet. Designed to help both managers and employees through the process, it covers everything from performance management philosophy to tactics, in the areas of goal setting, rating and evaluation, feedback and coaching, and development. Sections of the Web site are dedicated to explaining the business need and rationale for performance management, as well as for its behavioral model of performance. This company's appraisal process evaluates both progress toward goal attainment (the what) and competency (the how—or the behavior and skills important to success).

When it comes to rater training, this firm uses a combination of rater error training and frame-of-reference training. Embedded in the process is both structured diary keeping and structured recall. The training brings to raters' consciousness the possibility of various errors, but also explains that what looks like, for example, central tendency error may actually be an accurate rating. The company's Web site encouraged raters to be alert to the following errors:

- *Focusing on a Critical Incident*: Basing your assessment on one incident from the appraisal period and disregarding other performance. "Tina single-handedly managed the Chicago relocation."
- *Lower Rating for Less Challenge*: Rating some employees lower than others because they are in jobs that you believe are less challenging. "Why would anyone want to do administrative work? They must be low performers."
- *Nobody Can Be That Good*: Being an overly stringent rater, believing that no one is capable of "Very Effective" performance. "Samuel is by far the best employee that has ever worked for me, but he would have to walk on water before I give out a 'Very Effective' rating."
- *Similarity*: Giving high ratings to employees who are very similar to you. "Mark reminds me of when I started in technology. I've done well, so he must be a star."
- *Being Influenced by Prior Performance*: Believing that you must have overlooked something. "Pam received the highest rating last year; her performance seems a little off, so I must be overlooking something."
- *Rating for Retention*: Giving your employees higher ratings simply to keep them; ratings should be based on performance. "If I rate Joan low based on her results, she may leave; maybe I should rate her higher to keep her."
- *Style Differences*: Lowering an employee's rating because he or she approaches the work differently than you might. "Tony is too analytical. I can't really relate to him, and he doesn't approach things the way I do; I'm going to have to give him 'Partially Achieved Expected Results.'"

- *Emotional Rating*: Allowing strong feelings about an individual to influence a rating. For example, if you feel very positively about an employee, you may tend to see everything he or she does in a positive light. By the same token, if you have strong negative feelings, you may tend to see everything he or she does in a negative light.
- *Recent Performance Only*: Evaluating only the most recent events to determine a rating; a rating should be assessed on an entire year of performance.
- *Friendships*: Allowing personal relationships outside the office to influence an employee's rating. "John didn't perform up to par, but giving him the rating he deserves will damage our friendship, so I'll rate him 'Achieved Expected Results.'"

The human resource team works with line managers to help them further define and clarify what the different levels of expectations may look like for the various jobs or groups in their organizations.

Goal Setting

Organizations typically spend less time setting and establishing direction and goals than they spend evaluating employee performance at the end of the performance cycle. However, as organizations realize the value of having a robust performance management process, they are investing more time in establishing strategic direction and goal setting. The following are steps in a goal-setting process recommended by Mone and London (2002). While this is a logical sequence, day-to-day realities of organization life may not allow for completion of all the steps, particularly in the sequence we present. The first three steps are preliminary and a foundational part of good management practice, while the remaining steps are the core of a strong goal-setting program:

1. Managers tell their employees that they are starting the goal-setting process for the upcoming performance year and indicate the time frame in which they want to have the goal-setting process completed. Managers outline the process for them.
2. In preparation, managers advise their employees to do the following to create a context for current goal setting: re-read the mission and vision for the company; re-read the department's mission and vision; review their job descriptions; review their current goals, strategies, and tactics; identify any new overarching goals; and develop a working draft of team goals for their departments, if necessary (Fitzgerald, 1995). Of course, this assumes that these documents exist, are readily available, and are clear.
3. Managers share the overarching goals and department goals, if any, with their employees. They also share any other strategic messages that may be important for their employees to consider.
4. Using the above information, managers ask their employees to develop drafts of their performance goals, strategies, and tactics, and development goals.
5. Managers meet with each of their employees to review and discuss their goals, strategies, and tactics. During this meeting, managers will want to ensure that their employees' performance goals are aligned with the overall direction of the company and department, are challenging and meaningful, and are realistic. Managers will also want to ensure that each employee's development goals will help him or her to improve performance in the current job or acquire the skills or knowledge necessary to prepare for future jobs.
6. Once the goals, strategies, and tactics have been finalized to the managers' requirements, managers ask their employees to develop the goal measures of success and goal measurements.

7. Managers meet with their employees to review, discuss, and finalize measures.
8. Managers review their employees' goals, strategies, etc., with their managers, as necessary.
9. Managers communicate and discuss any changes with their employees and revise as necessary.
10. Managers and each of their employees should "sign off" on the agreed-to goals, strategies, etc., and each should keep a copy for their files.

When organizations provide employees with an abundance of tools and resources to empower them to control and manage their performance, they also equip and enable employees to act in a self-regulatory mode, which ultimately leads to greater focus on goal achievement and performance success.

Tracking and Evaluating Performance Management Programs

This article has addressed several critical gaps between research and practice in the management and assessment of employee performance, particularly with reference to rater training and participation in goal setting. Since performance management programs have such a significant effect on the employee and the organization, human resource practitioners need to ensure that their organization's system is grounded in the latest research, in step with best practices, and aligned with organizational goals. The most effective method for ensuring that an organization's performance management system possesses these characteristics is to implement an ongoing, systematic evaluation process.

The application of formal program evaluation techniques to human resource systems and processes is gaining increased attention (Edwards, Scott, & Raju, 2003) and is recommended as a powerful tool for human resource practitioners. A properly conducted evaluation not only addresses the impact that a performance management system has on bottom-line results, but also focuses on the effectiveness of its key design elements (e.g., goal setting, evaluation, rater training, feed-

back and coaching, etc.). Human resource practitioners who are familiar with how to properly evaluate performance management systems will be positioned to ensure that their process effectively addresses the changing needs of the organization and its employees. The extent to which an organization's performance management system supports accurate ratings and encourages effective goal setting can be determined through a systematic evaluation process. The results of this evaluation can be used to recommend practical and scientifically sound interventions for improvement.

A properly conducted evaluation not only addresses the impact that a performance management system has on bottom-line results, but also focuses on the effectiveness of its key design elements.

Evaluation of Rater Accuracy

Barnes-Farrell and Lynch (2003) have argued that an evaluation of rater accuracy should focus on the presence of conditions that support accurate ratings, such as trained, knowledgeable appraisers who understand the performance standards and have adequate opportunity to observe performance; have at their disposal a valid, reliable performance instrument; and operate in an organizational climate that facilitates and rewards accurate evaluations.

The first step in evaluating the supporting conditions for rater accuracy is to determine the appropriateness and relevance of the criteria by which employees' performances are being assessed. It is not unusual for organizations to revise or completely change their competency models every few years. While this may be sensible due to shifting corporate priorities, it can make the criteria for successful job performance a moving target. There are several key characteristics that need to be taken into account when evaluating performance criteria (Gottfredson, 1991). However, the predominant concern here is validity—to what extent are the performance criteria relevant to the organization's goals and individual's job requirements and how successfully have these criteria been translated into measures of performance—does the measure assess the performance it is designed to assess? An accurate assessment of performance can only be made when there is a clear link between the job requirements and the content of the evaluation.

Once the relevance of the performance criteria has been established, the other conditions supporting rater accuracy can be examined. The evaluation of the following three components of rater accuracy are addressed below: rating elevation and differentiation, performance rating instrument, and rater training.

Rating elevation and differentiation. As previously discussed, the evaluation of rater accuracy is challenging due to the absence of a "true" measure of performance to which actual performance ratings can be compared. The most important criteria for judging rater accuracy are those that most directly influence human resource decisions: (1) the overall elevation of performance ratings and (2) the degree to which ratings differentiate among employees, both within and across performance dimensions (Murphy & Balzer, 1989). Many organizations face the problem of inflated, undifferentiated performance ratings and, as a result, have difficulty relying upon this information for making critical personnel decisions. This has led some organizations to take such extreme measures as abandoning the use of performance ratings altogether or converting their performance measure into forced ranking systems. Both of these solutions can lead to undesirable consequences (e.g., excessive subjectivity, increased adverse impact) that could be avoided by implementing evaluation-derived recommendations that take into account both stakeholder needs and current research. While rater accuracy is challenging because there is no true score with which to compare ratings, this does not mean that "objective" performance measures are free of problems. Comparing ratings to objective measures such as bottom-line financial indicators, sales figures, or production figures (whichever might exist for the target job) may be useful; however, various deficiencies (e.g., inaccurate records, competition) and contamination (situational factors, such as peer performance) beyond individual performance may plague these measures as well.

The potential for rater bias against protected groups (e.g., women, ethnic minorities) must also be examined in the context of rating differentiation. That is, the performance ratings assigned to protected groups should be statistically compared to those assigned to the majority group to determine if there are significant differences. This analysis should be conducted on overall ratings as well as each performance dimension that factors into an overall rating. In addition, narrative documentation used to justify the ratings should be reviewed, with discrepancies noted between ratings for those employees receiving similar narratives. If a statistically significant difference is found in the ratings for one or more protected groups, it would be necessary to determine on what basis these differences exist and whether or not they are justified. This would include an investigation of demographic differences (e.g., education, job tenure), rating interactions between rater and ratee, and identification of other variables that may help explain these differences. A finding of performance differences should also lead to a review of the appraisal's validity, an evaluation of the raters' qualifications, and examination of the rater-training curriculum.

In addition to statistical analyses, focus groups, interviews, and reviews of relevant documentation are all useful tools for evaluating the existence and nature of rating elevation and differentiation. This analysis should be conducted on a yearly basis. As part of this review, it is generally recommended that any human resource systems that link to performance ratings, such as rewards and recognition provided across demographic groups, including compensation, promotion, bonus levels, and less formal rewards, be audited as well. The organization's legal staff should play an integral role in this part of the evaluation.

Performance-rating instrument. Barnes-Farrell and Lynch (2003) noted that there is no single, preferred rating instrument or format identified through best practices for ensuring accurate ratings. Their contention is that the evaluation of the measurement instrument hinges on two key questions: (1) Does the measure facilitate accurate communication of employee performance? and (2) Does its design align with the goals of measurement?

...the evaluation of rater accuracy is challenging due to the absence of a "true" measure of performance to which actual performance ratings can be compared.

The answers to these evaluation questions will have an impact on the instrument's accuracy, reliability, and acceptability of use.

Certain rating formats may facilitate communication of expectations and improve the quality of ratings (DeNisi, 1996). These formats are typically designed as behaviorally based instruments. This method involves standardizing the rating scale so that it is interpreted the same way by all raters and ratees. This is accomplished by attaching behavioral examples of performance, or "anchors," to various levels of the rating scale. These behavioral anchors are placed at points along the scale that reflect performance at varying proficiency levels. This serves to calibrate the raters and provide employees with examples of what is expected of their performance. The evaluation of rating formats should therefore answer the questions posed by Barnes-Farrell and Lynch (2003) and determine if the behavioral expectations are outlined to calibrate both rater and ratee.

Rater training. A key ingredient for ensuring accurate ratings is to provide training on all elements of the performance management process. Following the research cited in this article, the evaluation of the training curriculum should examine whether managers are trained on the performance philosophy of the organization and taught how to effectively gather information about performance expectations and make appropriate judgments. In addition, the training objectives should also be evaluated to ensure that managers understand and communicate the value of the performance management system and are fully capable of applying performance management standards uniformly when making judgments about employees.

The rater-training modules should be reviewed to ensure that frame-of-reference training has been included and if not, what alternative method is being used to help raters develop a performance schema for differentiating good from poor performance. The training evaluation should also determine if raters are being trained on the use of structured diary keeping and structured recall to help them organize performance information for later use in making ratings.

The curriculum should include training on the definitions and standards associated with the performance dimensions, the difference between observation and inference, documentation requirements, and a discussion of stereotypes/prejudices and their impact on ratings. In addition, raters should be trained on how to avoid errors in written comments, including comments that are inconsistent with numerical ratings.

Of course, human resource practitioners recognize that rater training is one part of a comprehensive appraisal system. Broader aspects of a performance management system are reflected in guidelines that incorporate rater training along with rating methods that are linked to important job behaviors and outcomes and use of appraisal results. For instance, Martin and Bartol (1998) recommended the following:

- Designate rating periods.
- Use clear and job-specific behaviorally based rating strategies, such as management by objectives (MBO) and behaviorally anchored rating scales (BARS).
- Base the appraisal on specific tasks the employee accomplishes or fails to accomplish, and the behaviors that are necessary to perform the job during the rating period.
- Provide space for written, individual standards of performance known to the employee.
- Conduct performance appraisal training for raters and ratees (training should include the role of the organization, rater, and ratee, and how the results will be used).
- Operate the system in a legally defensible manner (fair, job-related, bias-free).
- Link results of the performance appraisal process to the programs dependent on them (use the results to make decisions about people's pay, development, and assignments).
- Conduct performance appraisal on a schedule where findings support the organizational human resource systems.

A key ingredient for ensuring accurate ratings is to provide training on all elements of the performance management process.

- Clarify the standards used to appraise the job performance of the employees.

Finally, rater training should go beyond issues of improving rater accuracy to include providing constructive negative feedback. Managers generally dislike telling employees that they are not performing well. Moreover, employees don't want to hear negative messages and are likely to distort what is said and what they hear (cf. London, 2003). In addition, managers may not understand how to minimize employees' defensiveness. This is why coaches are often valuable, because they know how to provide such feedback. However, managers and other raters can be trained to provide this feedback.

Evaluation of Goal Setting and Participation

Participation in goal setting is important, but it must be accompanied by an understanding of organizational goals, information about translating them into action, and support for meeting goals.

Participation in goal setting is important, but it must be accompanied by an understanding of organizational goals, information about translating them into action, and support for meeting goals. The evaluation of individual goal setting and participation should begin with a review of the process that managers and employees follow throughout the organization to establish goals. As previously pointed out, participation in goal setting improves performance, not because participation by itself is inherently motivating, but because it provides the employee with an increased understanding of expectations and strategies for goal accomplishment. Therefore, regardless of whether goals are assigned or employees and managers work together to develop their goals, the concern from an evaluation standpoint is whether employees understand why their goals have been set and what are the specific expectations and strategies for goal accomplishment. This is not to say that participation in goal setting is an unimportant aspect of the evaluation—quite the contrary. As discussed earlier, participation in goal setting results in the establishment of more difficult goals, which in turn leads to increased commitment and higher performance. Therefore, an evaluation of how well an organization has incorporated learning about participative goal set-

ting into practice can help to identify useful areas for intervention that may directly impact its bottom line.

The evaluation of participation in goal setting should first address the extent to which the goal-setting process is formalized and consistently used across organizational levels. Key evaluation questions here include: Is participation in goal setting a cultural norm that extends all the way up to the highest levels? Do the manager and employee partner in setting goals and expectations and how is this executed and tracked? Are managers and employees trained on how to set actionable goals that are aligned with the organization's goals? Are goals calibrated across employees at the same organizational level and if so, how is this accomplished and what impact does this have on goal complexity and commitment? This calibration process may inadvertently water down goal complexity and commitment if there is an attempt to "standardize" goals across employees within a business unit. If goals are assigned, rather than participatively set, how is the information conveyed and do employees have a say in how these goals will be accomplished?

The evaluation of participation in the goal-setting process should also focus on the impact of changing organizational priorities. When organizational priorities shift, how rapidly is this communicated and what effect does this have on changing/updating goals? To what extent do employees participate in defining new goals under these conditions? Does the performance management process require and reinforce a joint effort and accountability between manager and employee to establish and update goals and to give and ask for feedback?

Interviews with key executives as well as focus groups and/or surveys of managers and employees are effective means for capturing how the system actually works and where gaps may exist. In addition, an analysis of the organization's policies and business rules concerning the goal-setting process should be included as part of this review.

Goal setting, coaching, and feedback. As noted above, feedback and goals work together to affect performance. Coaching,

from one's manager or a professional coach, supports feedback, participation in goal setting, and tracking goal accomplishment. Coaching and ongoing, timely, specific feedback play key roles in determining whether goals will lead to performance. As discussed above, coaching and feedback allow an employee to assess progress toward a goal and make necessary shifts in strategy as appropriate. Together, they reinforce the importance of the goal to the organization and help develop necessary proficiencies for goal accomplishment. An evaluation of the role of managers as coaches as an element of the performance management system will determine whether managers and employees meet regularly (at least quarterly) to discuss corporate goals and implications for their performance. Are these meetings tracked and documented? Are managers who demonstrate strong coaching skills recognized and rewarded? In addition, the evaluation should determine to what extent employees are actively involved in setting and seeking clarification for expectations, working toward goals, asking for feedback, keeping track of significant events, and identifying needs and resources. Is the process supported from the top down to enable open communication and the ongoing provision of feedback and coaching? The review of the coaching and feedback component should also identify what tools are available to managers and employees to assess barriers (individual and/or environmental) to goal accomplishment and what resources are at their disposal to address these issues. Evaluation data can be collected from both employees and managers through focus groups or surveys.

Training. The performance management training curriculum should be reviewed to ensure that managers and employees are trained on setting and writing goals that are specific, challenging, actionable, and aligned with the organization's goals. The training should also focus on increasing one's self-efficacy. Evaluation questions should address how well the training program teaches employees to operate in a self-regulatory mode and whether instruction is provided on techniques such as verbal self-guidance and

strategies for coping with barriers to goal accomplishment. In addition, the training should be evaluated to ensure that managers are trained on providing feedback, coaching/mentoring, and observing and documenting performance. Employee training should also be reviewed to ensure that employees are provided with information on the process and trained on how to work collaboratively with the manager to set goals and how to ask for and receive feedback.

Several studies examined attitudinal outcomes associated with goal-setting and performance management processes. Managers with a style that incorporates appraisal participation, goal setting, and feedback were viewed by employees as more effective counselors and communicators of feedback, and as having a more favorable impact on employee motivation and productivity (Reed & Roberts, 1996). Taylor and Pierce (1999) evaluated the introduction of a performance management system by examining changes in employees' satisfaction, ratings of cooperation with one's supervisor, and organizational commitment. They found that worker satisfaction and cooperation ratings increased for low performers but declined for high performers after the initial rating and bonus distribution. Presumably, high performers were put off by the introduction of the system.

Conclusion

We suggested ways that human resource managers can incorporate research results into the design and ongoing evaluation and refinement of ratings and the goal-setting component of performance management systems. Educating managers and employees in the range and meaning of the expected performance dimensions can help them recognize desired behaviors, evaluate performance accurately, and provide meaningful feedback, as well as guide their own goal setting and performance tracking. Appraisal discussions, feedback workshops, and coaching can facilitate these processes. The development of a performance management system is a continuous process reflective of changes in organizational direction and changes in technology (e.g., using online methods for training

Educating managers and employees in the range and meaning of the expected performance dimensions can help them recognize desired behaviors, evaluate performance accurately, and provide meaningful feedback, as well as guide their own goal setting and performance tracking.

and collecting performance ratings). Goal setting and performance evaluation can be used to establish a broad range of developmental assignments. Ongoing program evaluation is needed to identify areas for continuous improvement of the system.

Manuel London is a professor and director of the Center for Human Resource Management at the State University of New York at Stony Brook. He received his PhD in organizational psychology from Ohio State University. His areas of research are career and management development, multisource (360-degree) feedback, and organization change. Recent books include *Leadership Development* (2002, Erlbaum) and *Job Feedback* (2nd ed., 2003, Erlbaum).

Edward M. Mone has more than 25 years of experience in career, leadership, and organization change and development. He is currently vice president of organization development for Computer Associates, where he is responsible for learning and development, performance management, career management, and succession planning and leadership development. Previously, he held executive human resource positions at Cablevision Systems, Booz Allen & Hamilton, and AT&T. He is an adjunct professor of management at the State University of New York at Stony Brook. He holds an MA in counseling psychology. His recent books are *HR to the Rescue: Case Studies of HR Solutions to Business Challenges* (1998, Gulf) and *Fundamentals of Performance Management* (2002, Spiro Business Guides).

John C. Scott is vice president and cofounder of Applied Psychological Techniques (APT), a human resources consulting firm that specializes in the design and validation of selection and assessment technologies, staffing for organizational change, performance management, and employment litigation support. He received his PhD from the Illinois Institute of Technology in 1985. He directs consulting services in the areas of selection development and validation, skills assessment, survey design, performance management, and executive assessment. Scott has served on the program and review committees for Division 14 of the American Psychological Association, the Society for Industrial and Organizational Psychology's Professional Practice Book Series editorial board, and is a frequent presenter in the area of selection and assessment. He is coeditor of *The Human Resources Program-Evaluation Handbook* (2003, Sage).

REFERENCES

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: Freeman.

Barnes-Farrell, J.L, & Lynch, A.M. (2003). Performance appraisal and feedback programs. In J. E. Edwards, J. C. Scott, & N. S. Raju (Eds.), The human resources program-evaluation handbook (pp. 155–176). Thousand Oaks, CA: Sage Publications.

Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. Academy of Management Review, 6, 205–212.

Cronbach, L. J. (1955). Processes affecting scores on understanding of others' and "assumed similarity." Psychological Bulletin, 59, 177–193.

DeNisi, A. S. (1996). Cognitive approach to performance appraisal: A program of research. New York: Routledge.

DeNisi, A. S., & Peters, L. H. (1996). Organization of information in memory and the performance appraisal process: Evidence from the field. Journal of Applied Psychology, 81(6), 717–737.

Drucker, P. (1954). The practice of management. New York: Harper.

Edwards, J.E., Scott, J.C., & Raju, N.S. (2003). The human resources program-evaluation handbook. Thousand Oaks, CA: Sage Publications

Fitzgerald, W. (1995). Forget the form in performance appraisals. HR Magazine, 40(12), 136+.

Gottfredson, L.S. (1991). The evaluation of alternative measures of job performance. In A. K. Wigdor, B. F. Green Jr. (Eds.), Performance assessment for the workplace, Volume 2: Technical issues (pp. 75–126). Washington, DC: National Academy of Sciences Press.

Greguras, G. J., Robie, C., Schleicher, D. J., & Goff, M., III. (2003). A field study of the effects of rating purpose on the quality of multisource ratings. Personnel Psychology, 56, 1–21.

Hakel, M. D. (1980, April). An appraisal of performance appraisal: Sniping with a shotgun. Paper presented at the First Annual Scientist-Practitioner Conference in Industrial-Organizational Psychology, Virginia Beach, VA.

Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. Journal of Applied Psychology, 73(1), 68–73.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64, 349–371.

Klein, H. J., Wesson, M. J., Hollenbeck, J. R., & Alge, B. J. (1999). Goal commitment and the goal-setting process: Conceptual clarification and empirical synthesis. Journal of Applied Psychology, 84(6), 885–896.

Latham, G. (2002). The reciprocal effects of science on practice: insights from the practice and science of goal setting. Canadian Psychology, 42, 1–11.

Latham, G. P., Erez, M., & Locke, E. (1988). Resolving scientific disputes by the joint design of crucial experiments by the antagonists: Application to the Erez-Latham dispute regarding participation in goal setting. Journal of Applied Psychology, 73, 753–772.

Latham, G. P., & Marshall, H. A. (1982). The effects of self-set, participatively set, and assigned goals on the performance of government employees. Personnel Psychology, 35, 399–404.

Latham, G. P., Mitchell, T. R., & Dosset, D. L. (1978). The importance of participative goal setting and anticipated rewards on goal difficulty and job performance. Journal of Applied Psychology, 63, 163–171.

Latham, G. P., & Saari, L. M. (1979). The effects of holding goal difficulty constant on assigned and participatively set goals. Academy of Management Journal, 22, 163–168.

Latham, G. P., Steele, T. P., & Saari, L. M. (1982). The effects of participation and goal difficulty on performance. Personnel Psychology, 35, 677–686.

Latham, G. P., & Yukl, G. A. (1976). Effects of assigned and participative goal setting on performance and job satisfaction. Journal of Applied Psychology, 61, 166–171.

Locke, E. A. (1991). The motivation sequence, the motivation hub, and the motivation core. Organizational Behavior and Human Decision Processes, 50, 288–299.

Locke, E. A., Alavi, M., & Wagner, J. (1997). Participation in decision making: An information exchange perspective. In G. Ferris (Ed.), Research in personnel and human resources management (Vol. 15, pp. 293–331). Greenwich, CT: JAI Press.

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting and task performance. Englewood Cliffs, NJ: Prentice Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. American Psychologist, 57(9), 705–717.

London, M. (2003). Job feedback (2nd ed.). Mahwah, NJ: Erlbaum.

Martin, D. D., & Bartol, K. M. (1998). Performance appraisal: Maintaining system effectiveness. Public Personnel Management, 27(2), 223–230.

Mone, E. M., & London, M. (2002). Fundamentals of performance management. London, England: Spiro.

Murphy, K. R., & Balzer, W. K. (1989). Rater errors and rating accuracy. Journal of Applied Psychology, 74(4), 619–624.

Oettingen, G., Pak, H. J., & Schnetter, K. (2001). Self-regulation of goal setting: Turning free fantasies about the future into binding goals. Journal of Personality and Social Psychology, 80(5), 736–753.

Reed, T., & Roberts, G. E. (1996). Performance appraisal participation, goal setting, and feedback. Review of Public Personnel Administration, 16(4), 29–60.

Rodgers, R., & Hunter, J. E. (1991). Impact of management by objectives on organizational productivity. Journal of Applied Psychology, 76(2), 322–336.

Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. Human Resource Management, 41, 149–174.

Stamoulis, D. T., & Hauenstein, N.M.A. (1993). Rater training and rating accuracy: Training for dimensional accuracy versus training for ratee differentiation. Journal of Applied Psychology, 78(6), 994–1003.

Sulsky, L. M., & Balzer, W. K. (1988). The meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. Journal of Applied Psychology, 73, 497–506.

Taylor, P. J., & Pierce, J. L. (1999). Effects of introducing a performance management system on employees' subsequent attitudes and effort. Public Personnel Management, 28(3), 423–452.

Woehr, D. (1994). Understanding frame-of-reference training: The impact of training on the recall of performance information. Journal of Applied Psychology, 79(4), 525–534.

[zaq1:[**AUTHOR: Should this be "performance of expected."?]]**

<zaq2>[[AUTHOR: I think this would make more sense if it read "performance, goals and self-efficacy will be more highly related when goals are self-set." Please advise.]]