Chapter 7

ADVANCES IN THE SCIENCE OF PERFORMANCE APPRAISAL: IMPLICATIONS FOR PRACTICE

Gary P. Latham and Sara Mann
Rotman School of Management, University of Toronto, Canada

There have been at least five reviews of the literature on performance appraisal. Two appeared in this series 20 years ago (Latham, 1986) and 13 years ago (Latham, Skarlicki, Irvine, & Siegel, 1993). A third review was conducted five years later by Arvey and Murphy (1998). The fourth and fifth interrelated reviews by Fletcher (2001) and Fletcher and Perry (2001) appeared five years ago.

The topics focused on in the first two time periods are re-examined in the present chapter. They include legal issues affecting performance appraisals, performance criteria, and the scales that can be used for assessing people on them, possible sources of an appraisal, objectivity/accuracy of appraisers, and ways of bringing about and sustaining a person's performance effectiveness. The literature on legal issues in the first two time periods was restricted primarily to North American organizations. This is no longer the case in this new millennium. The present chapter describes legislation affecting performance appraisals conducted in Asia, Australia, Europe, and Latin America. Contrary to the conclusions reached in the two earlier reviews, there is now evidence that appraisers do discriminate against people on the basis of demographic variables. As a result of Bommer, Johnson, Rich, Podsakoff, and MacKenzie's (1995) meta-analysis revealing the convergent validity of hard versus soft criteria, the debate around the use of these two performance measures in the first two time periods has ended. The current focus is on the identification and measurement of employee behaviors, particularly those that are task related, which impact an organization's effectiveness. The focus on different methods of job analysis that occupied the attention of researchers in the 1980s disappeared in the 1990s and has not been resurrected. This domain has been mastered. User reactions to appraisals were only touched upon in the 1980s–1990s,

when the emphasis was on psychometric concerns regarding the scales used for making appraisal decisions. The converse is true in the present time period. We know how to build reliable and content valid appraisal scales. We are currently discovering how to get people to like them so much that they will use them. This is necessitated by the fact that many managers are so uncomfortable conducting appraisals that they give uniformly high ratings to their employees (Tziner & Murphy, 1999). The employees who are on the receiving end of the appraisal express dissatisfaction with both the decisions made as a result of a performance assessment and the process of performance assessment. As Bowman (1999, p. 557) wryly observed, a performance appraisal is 'given by someone who does not want to give it to someone who does not want to get it.'

With regard to sources of appraisal, the primary focus of research in the 1980s was on supervisory, peer, and self-appraisals. In the 1990s, upward feedback from subordinates was also being studied. No mention was made in either time period on the use of multisources commonly referred to as 360-degree feedback. Research on this latter topic has subsequently blossomed in North America and the United Kingdom (Fletcher, 2001). Research on cognitive information processes as a way to improve appraiser accuracy, relative to the two earlier time periods, has decreased significantly. There is general agreement that this research stream was not productive (e.g., Arvey & Murphy, 1998; Fletcher, 2001). As Fletcher and Perry (2001) noted, the findings from the voluminous laboratory experiments on this subject did not take into account the motives of appraisers, and how these motives affect appraisal decisions. With regard to the theoretical frameworks guiding appraisal research, attribution theory is no longer used. Researchers continue to draw upon goal setting and organizational justice theories. Finally, the perennial debate on whether the developmental aspects of a performance appraisal should be temporally separated from those that are evaluative has subsided. This is the result of a monumental shift in the attention of researchers from the discrete performance appraisal to ongoing performance management where coaching is inherent in the process. Thus this may be the last review of the literature where performance appraisal is in the title. The present chapter concludes with a research agenda for performance management based on findings from ongoing research in the motivation and decision-making fields of industrial and organizational (I/O) psychology as well as those in clinical/counseling psychology. These issues have not been discussed in previous literature reviews.

LEGAL ISSUES

Because organizations have become increasingly global in the twenty-first century, it has become important for organizational decision makers to be aware of and sensitive to legislation and court decisions governing performance appraisal, both of which reflect a country's culture. In doing so, Latham and

McCauley (2005) predicted that, as this new millennium progresses, there will be a shift in emphasis from the twentieth-century focus on diversity issues within a country regarding an employee's race, sex, or age to organizational leaders having to take into account differences among countries regarding employee ethnicity, national origin, and political ideology. For example, Fletcher and Perry (2001) observed that the purpose of performance appraisals varies across cultures. In individualistic, low-power cultures typified by Canada and the USA, appraisals are often used to differentiate among employees for making pay and promotion decisions. Employees expect and are expected to participate in the appraisal process. In collectivistic, high-power cultures typified by many countries in Asia and Latin America, the developmental aspects of appraisals are emphasized relative to the evaluative so as to increase organizational commitment. Employees neither expect nor are they expected to be involved in this process. Hence, performance appraisals are likely to be most effective and accepted when a country's cultural values are taken into account in their design and implementation (Erez, 1994). But, despite the large shadow that the law casts on this domain, there is an absence of empirical research on the role that cultural differences actually play in performance appraisals (Fletcher & Perry, 2001).

The need for this research is evidenced by the fact that people in the United States, a culturally diverse society, have become increasingly aware of their legal rights. Martin, Bartol, and Kehoe (2000) reported a 100% increase in the number of discrimination cases filed in 1999 over the number filed in 1995. Moreover, these court cases usually involved complaints regarding a performance appraisal (Arvey & Murphy, 1998). Werner and Bolino (1997) updated the findings of Feild and Holley (1982) on ways that organizations are likely to win a court case. Emphasis is given by the US courts to evidence of fairness and due process—subjects that are discussed in more detail in a subsequent section. Issues that are of lesser importance to the court include the type of appraisal scale and evidence of validity.

Legal concerns pertaining to performance appraisal are by no means restricted to North America (Woodford & Maes, 2002). For example, Head, Haug, Krabbenhoft, and Ma (2000) reported that in Latin America, countries such as Panama, Nicaragua, Bolivia, and Argentina have outlawed discrimination based on political opinion. However, in Chile, Guatemala, and Paraguay no groups are protected and employment discrimination can occur openly. In Asia, India prohibits discrimination on the grounds of religion, race, caste, sex, and place of birth or residence (Jain, Sloan, & Horowitz, 2003).

Western-oriented countries have passed legislation that is more similar than different from the USA. For example, Australia has comprehensive anti-discrimination legislation that prohibits the use of gender, sexual preference, age, marital status, child or elder care responsibilities, race, religion, political belief, pregnancy, physical features, industrial activity, and impairment in any and all aspects of employment decision making, including recruitment,

selection, performance appraisal, training, compensation, and separation (Bennington & Roberts-Calvert, 1998; Bennington & Wein, 2000). As of 1994, legislation in South Africa provides equal opportunity to all citizens regardless of color, gender, religion, political opinion, or sexual orientation (Jain et al., 2003). In 1996, legislation in Northern Ireland was passed prohibiting employment discrimination on the basis of religion, political opinion, color, nationality, race, and ethnic or national origin (Jain et al., 2003). Similarly, the Race Relations Act in Britain outlaws direct and indirect discrimination on the grounds of race, color, and nationality, including citizenship or ethnic or national origins, but not on religious grounds (Jain et al., 2003). The European Union (EU) has not passed new legislation affecting performance appraisals. Instead the EU has kept in place the local labor laws of the 15 member countries (Claus, 2003).

APPRAISER BIAS

Is the concern of legislators and the courts justified regarding the fairness of an employee's performance appraisal? The empirical literature suggests that the answer is 'Yes.'

Lance (1994) found that 'ratings were stronger reflections of raters' overall biases than of true performance factors' (p. 768). A massive study involving over 4000 managers where there were appraisals from at least two supervisors, two peers, two subordinates, plus self-appraisals, indicated that individual rater effects account for over half of the rating variance (Scullen, Mount, & Goff, 2000).

A field study involving the US Army showed that the person's knowledge and ability explained only 13% of the variance in an appraisal from a supervisor, and just 7% in appraisals from peers. Future research should focus on the extent to which this is due to supervisors and peers weighting issues of interpersonal facilitation skills and trust as more important than a person's current knowledge and ability. If this is true, the appraisals reflect important variance on the relationship aspects of a person's job. Support for this hypothesis can be inferred from the research on the multidimensions of performance discussed subsequently in this chapter.

A review of the literature revealed that the supervisor's positive regard for a subordinate resulted in leniency and halo errors, and less inclination to punish poor performance (Lefkowitz, 2000). Studies in the private sector have found that the perceived similarity of the subordinate by the supervisor inflates the appraisal (Wayne & Liden, 1995). This is especially true with regard to perceived similarities regarding extroversion, conscientiousness, and emotional stability (Strauss, Barrick, & Connerley, 2001). In short, to the extent that appraisers viewed an employee as similar to themselves, they tended to like the person more, and consequently, gave the person a higher rating. Again, research is

needed on the extent to which appraisers are focusing on the interpersonal facilitation skills required of a person to be effective in today's team-driven organizations rather than on traditional performance metrics (e.g., goals set vs goals attained) that are often affected by factors beyond a person's control (e.g., the economy).

A person's appraisal is also affected by gender, a finding that does not go unnoticed by the courts. A review of objective and subjective indicators of performance as a leader revealed that men are usually evaluated as more effective than women (Eagly, Karau, & Makhijani, 1995). That this finding, at least in part, reflects discrimination is suggested by the fact that even when the sexes engaged in equivalent leadership behavior, women were devalued when leadership roles required a 'masculine' leadership style, and the appraisers were male (Eagly, Makhijani, & Klonsky, 1992).

A subsequent review of leadership perceptions in the military revealed preferential ratings for men in training groups primarily where there was a 'token' woman; this typically was not found in groups where there were several women (Biernat, Crandall, Young, Kobrynowicz, & Halpin, 1998). Token status exacerbates less positive evaluations of women than men, because tokens receive considerable attention that increases the pressure on them to perform well (Biernat et al., 1998).

Boldry, Wood, and Kashy (2001) found that sex proportion is a factor affecting rating accuracy. The respondents, only 27 of whom were female, reported that men possess the motivation and leadership qualities necessary for effective performance, whereas women possess feminine attitudes that impair effective performance. Yet there were no sex differences on any performance measure (i.e., grades, rank, physical training scores). The authors concluded: 'Overall, these results indicate that sex differences in evaluations do not reflect a kernel of truth as represented by sex differences in performance or by sex differences in self ratings' (p. 702). These findings link the lower evaluations of women to general gender stereotypes and show that these stereotypes extend to perceived motivation as well as leadership.

In a study of over 3000 high-ranking military officers (majors, lieutenant-colonels) in Israel, five variables were analyzed, namely, overall performance, advancement prospects, analytic competence, operational competence, and interpersonal relations (Pazy & Oron, 2001). The results showed that the performance of women was rated higher than that of men when the women constituted a higher proportion than the men. Sex proportion appears to be a contextual variable of importance in explaining gender-related influences on performance evaluations (Bartol, 1999). Increasing the representation of women in mixed groups increases their positive appraisals. The authors concluded that under-representation of women in military units, even at very high ranks, has a deleterious effect on the performance appraisals they receive. These findings are similar to those obtained by Kraiger and Ford (1985) regarding the evaluations of African-Americans.

In a field study of 1268 managerial employees, Landau (1995) found that controlling for age, education, tenure, salary grade, functional area, and satisfaction with support for one's career, both race and gender were related to appraisals of one's promotion potential. Females were rated lower than males, and blacks and Asians were rated lower than whites. Just as it has been found that both black and white appraisers assessed the performance of black employees as lower than that of white employees, Dewberry (2001) reported similar effects for ethnic minority vs white comparisons in the United Kingdom.

A meta-analysis by Bowen, Swim, and Jacobs (2000) provides a ray of optimism with regard to these findings. They found no evidence of gender bias in performance ratings when all the raters are not male, and when the appraisal instrument is not gender-typed, that is, contains no male or female stereotypes of performance.

In a field study of entry level employees in accounting firms, Saks and Waldman (1998) found that when work experience, in terms of the number of previously held jobs and a person's grade point average, is controlled, there is no relationship between age and performance evaluations. Gordon and Arvey (2004) conducted a meta-analytic review of age discrimination research on job incumbents that provides still another ray of hope in this area. Their results suggest that age bias is less of a problem today than it was in previous decades. The greater and more relevant the information that is provided regarding an employee, and the greater and more relevant the experience of the raters, the less evidence of age bias.

Gilbert, Hannan, and Lowe (1998) found that people who smoke are rated lower than non-smokers on professional comportment, working with others, and dependability. This is likely the result of smokers losing production time as a result of leaving for designated smoking areas, and due to rater bias against those who smoke.

A person's reputation also plays a role in decisions regarding rewards that are to be administered. Johnson, Erez, Kiker, and Motowidlo (2002) found that a helpful person who has a good reputation received more rewards than did a helpful person who has a bad reputation. An unhelpful person, with a good reputation, did not receive more rewards than an unhelpful person with a bad reputation. These findings suggest that when a person is unhelpful, a good reputation does not help. But when a person's performance is helpful, a bad reputation continues to adversely affect the rewards that are subsequently administered despite the person's good performance.

The motivation attributed by a supervisor to an employee also affects appraisals of that person's performance. In a laboratory experiment, Pelletier and Vallerand (1996) found that even when there are no differences in job performance among subordinates, supervisors evaluate an employee's performance more positively when they believe that the subordinate is motivated to perform the task for intrinsic rather than extrinsic reasons. In short, a performance appraisal is influenced not only by the performance demonstrated by an

employee, but also by what a supervisor perceives the motivation of the employee to be.

DeVoe and Iyengar (2003) also examined the relationship, in three different cultures, between a manager's perception of an employee's motivation and the subsequent performance appraisal. North American managers perceived their employees as being more extrinsically than intrinsically motivated. Perceptions of intrinsic motivation resulted in a high performance appraisal. Asian managers perceived their subordinates as being motivated by both factors, and their perceptions affected performance appraisals positively. Latin American managers perceived their employees as being more intrinsically motivated, and only their perceptions of an employee's intrinsic motivation correlated significantly with an employee's performance appraisal.

In short, legislation and court decisions alone are not eliminating bias in performance appraisals. This does not bode well for organizations whose leaders must find ways to minimize rigidity, insensitivity in interactions, and intolerance among different ethnic groups (Latham & McCauley, 2005). The solutions to this problem may be at least three-fold, namely, find ways to increase user acceptance of the appraisal process, base appraisal decisions on multi-sources, and train appraisers not only to be objective, but to coach employees on a year-round basis. Research is now needed to test the effect of this treatment intervention on perceived fairness in performance evaluations. Advances that have been made on the individual components of this proposed 'treatment package' are reviewed below.

THE APPRAISAL SCALE AND USER REACTIONS

An organization's strategic plan often fails because little or no attention is given to how it should be operationalized behaviorally in the field (Latham & Latham, 2000). An appraisal scale that specifies what an employee must start doing, stop doing, or continue doing becomes a diagnostic instrument that facilitates self-management, as well as coaching by others of an employee. Such an instrument becomes highly relevant for the employee and the appraiser(s) because it specifies the behaviors necessary for the strategic plan to succeed. As a result, such an instrument is difficult to attack in a court of law. Hence the importance of job analysis for identifying critical behaviors cannot be overemphasized. Cronshaw (1998) argued the merits of a future-oriented analysis where the emphasis is on the anticipation of behaviors that will become necessary for the effective implementation of an organization's strategy.

Failure to develop appraisal scales that operationalize behaviorally what must be done to implement an organization's strategy probably explains Fletcher's (2001) finding. He reported that most organizations in the United Kingdom remain dissatisfied with their appraisal systems, not only because these systems cannot be shown to be psychometrically valid, but because they are not

perceived as a means for developing and motivating people. Hence legal challenges initiated by employees to an appraisal that adversely affects them should not be surprising.

In short, a primary reason for the frequent failure of a performance appraisal to bring about a positive change in a person's behavior is that the appraisal instrument often measures the 'wrong things.' In Australia, a study of more than 400 Army officers revealed that a substantial proportion reported concerns regarding the utility of the appraisal instrument as a device for accurately assessing work performance and promotion potential (Salter, 1996).

Job analysis methods for correcting the issues described above were developed in the past century. The emphasis now needs to shift from science to practice regarding the measurement of an employee's performance. Hence little attention is currently being given by researchers to the development of additional job analysis techniques.

Attention continues to be given to what constitutes performance effectiveness. Conway (1996, 1999), using factor analysis and a multitrait–multirater database, found that employee performance should be defined as either task or contextual. The latter is often referred to as organizational citizenship behavior or OCB (Organ, 1997). Van Scotter and Motowidlo (1996) found that OCB can be further dichotomized in terms of interpersonal facilitation and job dedication. However, from a practical standpoint, OCB can be treated as an aggregate construct. A meta-analysis revealed that the two dimensions are not only highly correlated, but that there are no meaningful differences in their relationships with such measures as job satisfaction, organizational commitment, or conscientiousness (LePine, Erez, & Johnson, 2002). Rotundo and Sackett (2002) found that job performance has three dimensions, namely, task performance, OCB, and counterproductive performance. An appraisal instrument should assess people on each of these three factors in relation to an organization's strategy.

Similar to the decrease in research emphasis on the development of new methods of job analysis, there has been a decrease in emphasis by researchers on the development of new appraisal instruments. An exception to this statement is an innovative approach to the development of an appraisal instrument that was tested in a laboratory experiment by Borman et al. (2001). They found that a computerized adaptive rating scale (CARS) had higher interobserver reliability and accuracy than a behaviorally anchored rating scale (BARS). In brief, this appraisal format presents the appraiser with two behavioral statements associated with a given dimension of performance, one reflecting 'below average' and the other 'above average' performance. Depending on which of two statements the appraiser chooses as describing the ratee, the computer selects two additional behavioral statements; one with a scaled effectiveness level above, and one below the effectiveness value of the statement chosen initially. A ratee's 'true' level of effectiveness is determined by these iterative paired comparisons. User reactions in field settings to CARS have yet to be tested.

There is evidence suggesting that the number of categories, three vs five, that an organization allows for describing a person's performance affects motivation and perceptions of fairness. Bartol, Durham, and Poon (2001), in a laboratory experiment, found that the number of performance categories was positively related to (a) the person's self-efficacy for attaining the next higher rating category, (b) higher goals being set, and (c) higher performance, with five being superior to three.

Jack Welch, the former CEO of the General Electric Company, argued that: 'A company that bets its future on its people must remove the lower 10%, and keep removing it every year—always raising the bar of performance . . .' (General Electric Company, 2000, p. 4). Through the use of a simulation, Scullen, Bergey, and Aiman-Smith (2005) investigated the effect of this forced distribution rating system on the performance of a workforce. They found that there is indeed a noticeable improvement in the first several years. However, their findings also show that eventually each time a company improves its workforce by replacing an employee with a new hire, it becomes more difficult to do so again. This is because an organization eventually reaches the point where its poorest employees are as capable as its best applicants. Employees at Microsoft, Conoco, Ford, Goodyear, and General Motors have filed class-action law suits in the USA claiming that this appraisal system resulted in discrimination against blacks, women, older employees, and non-citizens (Scullen et al., 2005).

With the emphasis that the courts place on the concept of fairness in appraising employees, Sulsky and Keown (1998) concluded, after reviewing the appraisal literature, that research emphasis needs to be placed on ways to improve user reactions to the appraisal process. Mayer and Davis (1999), in a longitudinal field study, found that trust in top management increased significantly when a new performance appraisal system was introduced to replace one that was viewed by the employees as inaccurate and not allowing for performance-related rewards. Similarly, in an empirical study involving the US Air Force's use of BARS, Hedge and Teachout (2000) argued that psychometric considerations regarding the reliability and validity of appraisals will always need to be taken into account, but they are not sufficient criteria for evaluating an appraisal system. Critical to the implementation and ongoing use of appraisals is user acceptance. Trust, as well as perceptions of situational constraints, predicted the motivation to rate others, as well as the motivation to rate them accurately. Trust was defined as perceptions of whether people adhere to the rules when making appraisals, whether the appraisals are accurate, and whether the appraisals reflect favoritism. Situational constraints include the absence, or lack of clarity, of a technical manual for conducting performance appraisals.

A meta-analysis by Cawley, Keeping, and Levy (1998) revealed that the employee's motivation to improve performance, employee satisfaction with the appraisal system as a whole, as well as with the appraisal session itself, is high if the person is allowed to participate in the appraisal process. Participation

that results in perceptions by employees that their viewpoints were taken into account in administrative decisions has a greater impact on their satisfaction than participation for the purpose of influencing the end result. In short, allowing employees 'voice' influences their perceptions of fairness. Consistent with these findings, Kleingeld, Van Tuijl, and Algera (2004), in a field study involving employees of a Dutch supplier of office equipment, found that implementing a performance management system that involves employee participation can lead to a significant increase in performance. In addition, they found that employee satisfaction with the program, and the perceived usefulness of the feedback, were significantly higher when the performance appraisal system included employee participation.

In their review of the literature on user reactions to different appraisal instruments, Tziner and Kopelman (2002) reported that an appraisal system is often abandoned if it fails to elicit positive reactions from the user. Behavioral Observation Scales or BOS (Latham & Wexley, 1994) were found to be superior to other appraisal instruments in terms of eliciting favorable attitudes. They concluded that this is because, relative to other instruments such as BARS and trait scales, BOS convey precisely what an individual 'must do,' feedback is perceived as factual, objective, and unbiased; and BOS feedback is conducive to setting specific high goals that are job related. Users preferred trait scales over BARS (see also Bernardin, 2005). In addition, Varma, DeNisi, and Peters (1996) found that bias in ratings is reduced when the appraisal instrument requires appraisals based on observable behavior. The bias of the appraiser, as noted earlier, continues to plague appraisal decisions.

SOURCES OF APPRAISAL

A critical component of any appraisal system is the appraiser (Sulsky & Keown, 1998). Thus, once the decision has been made as to the appraisal instrument that is acceptable to stakeholders, the issue remains as to who should use it. Who should do the appraisal? Is the answer supervisors, peers, subordinates, the employees themselves, or all of the above?

Supervisory Appraisals

Walker and Smither (1999) found that a subordinate's performance increases significantly more in those years in which the supervisor holds appraisal meetings than in the years where such meetings do not take place. This finding indicates that in spite of the biases that can enter into an appraisal, sitting with and talking to an individual is a worthwhile process for bringing about a subsequent improvement in a person's performance.

There are data that suggest that supervisors spend less than 1% of their time observing their subordinates (Komaki & Desselles, 1994). Appraisal errors

are highly likely when an observer has limited opportunity to observe an employee. Consequently, another reason why appraisals can fail to improve a person's performance is employee hostility toward the supervisor. Difficulty in answering satisfactorily such questions as: 'On what basis are you able to evaluate me?' undermines the credibility of the source.

Hence it would appear that supervisors, rather than being the primary source for observing an employee's performance, should be held accountable primarily for collecting data for appraising an individual from multiple sources, and then making the final appraisal decision based on these multiple sources of information. Administrative and developmental decisions based on multiple sources are likely to have high credibility in the courtroom as well as in the court of public opinion. Moreover, multiple sources of feedback increase the probability that the person will take positive action as a result of the feedback (Smither, London, Flautt, Vargas, & Kucine, 2003). Sources of an appraisal, in addition to supervisors, include an employee's peers, subordinates, as well as self.

Peer Appraisals

Greguras and Robie (1998) found that peers are the most reliable source of performance information followed by subordinates and superiors, respectively. They concluded that 'although reliability does not necessitate validity, these results suggest that researchers and practitioners may be well advised to seek, and perhaps weight more heavily information from peers and subordinates' (p. 964).

In a laboratory experiment involving the use of BOS, anonymous peer appraisals increased a person's interpersonal effectiveness (Dominick, Reilly, & McGourty, 1997). In a study involving MBA students working in project teams, peer appraisals increased group cohesion, openness of communication, task motivation, and group satisfaction (Drusket & Wolff, 1999). These findings are consistent with Conway's (1999) study of managers. Peers tend to give more emphasis to interpersonal skills than to task performance in making overall performance ratings, whereas the converse is true for supervisors. It was found that a subsequent improvement in performance resulting from peer feedback is especially likely if the person scores highly on conscientiousness and openness to experience.

In a study of the Special Forces in the US Army, peer rankings predicted final training outcomes better than did ratings from the staff. Again, it was found that peers placed significantly more importance on interpersonal performance and motivation than did the staff, and they placed significantly less importance on task performance (Zazanis, Zaccaco, & Kilcullen, 2001). Fletcher (2001) concluded that peers are the optimal source for assessments of a person's OCB.

Anonymous peer appraisals are becoming increasingly accepted as teams are becoming critical to the way that work is accomplished (Cohen & Bailey, 1997).

Moreover, teams are expected to be self-managing groups (Druskat & Wolff, 1999). Thus the people who comprise the teams are responsible for appraising and coaching one another.

Bettenhausen and Fedor (1997) reported that one's relationship with one's supervisor has little effect on one's perspective regarding peer appraisals. In contrast, only those with good coworker relationships supported their use, even when they are used for developmental rather than administrative purposes.

Among the reasons peer appraisals have been found to be unacceptable in previous studies is that inadequate appraisal instruments were used, namely, global trait-like measures. A robust finding in the behavioral sciences is that negative outcomes are perceived as particularly unfair when they are not adequately explained; they are perceived as fair when they are accompanied by an informative explanation (Greenberg, 2000). Consequently, Latham and Seijts (1997) looked at whether the type of appraisal instrument, namely, BOS, BARS, trait scales, or a blank sheet of paper that is used to provide feedback to managers, affects satisfaction with a peer appraisal. The results supported the hypothesis that peer appraisals are viewed as most useful and most fair when based on either BOS or a blank sheet of paper. Use of the latter, however, is likely to be problematic in a court of law as it invites idiosyncratic decisions on the part of appraisers that may not be job related.

Subordinate Appraisals

Atwater, Roush, and Fischthal (1995) found that anonymous feedback from subordinates promotes positive changes in a leader's behavior. Moreover, the study showed that leaders who receive feedback from subordinates that is more negative than their own self-evaluation show the greatest level of subsequent improvement.

Smither et al. (1995) obtained similar results in a study involving managers who were assessed on BOS. The greatest improvement was observed on the part of those managers whose self-evaluation was initially higher than the evaluation they received from their subordinates. These changes in behavior on the part of managers were not only positive, but they were shown to be sustainable over time (Reilly, Smither, & Vasilopoulos, 1996; Walker & Smither, 1999).

Heslin and Latham (2004), in a study of an Australian professional services firm, found that upward feedback from subordinates improves the performance of managers. Two key moderators are the manager's self-efficacy, that is, the 'belief that I can change,' as well as a learning goal orientation, that is, a focus on performance improvement rather than 'one's score.'

As is the case with peers, subordinate evaluations must be anonymous. Antonioni (1994) found, not surprisingly, that subordinates who were not allowed anonymity viewed the process more negatively than those who were anonymous, so much so that many non-anonymous raters refused to provide appraisals. Those subordinates who continued to give appraisals were significantly more lenient in their appraisals than their anonymous counterparts.

The purpose of an appraisal affects the assessments given by subordinates. Greguras, Robie, Schleicher, and Goff (2003) found that the quality of subordinate ratings is significantly better when they are to be used for developmental rather than for administrative purposes. However, Smither and Walker (2004) found that people who receive more favorable narrative comments from their direct reports also received a more favorable annual review from their supervisor. And Ostroff, Atwater, and Feinberg (2004) found that subordinate, in addition to peer, evaluations were related to a supervisor's overall evaluation of an employee. This may reflect the fact that supervisors become knowledgeable of a person's performance by listening to what others have observed.

Self-appraisals

That agreement between a self-appraisal and appraisals from other sources is usually lacking was shown in a meta-analysis by Beehr, Ivanitskaya, Hansen, Erofeev, and Gudanowski (2001) of over 2000 employees who received supervisor, peer, and self-appraisals. Moreover, self-appraisals had the least predictive validity regarding a person's future performance.

Although employees generally rate their performance more favorably than do their supervisors, Korsgaard, Meglino, and Lester (2004), in a study of employees at a healthcare claims center in the Midwestern US, found that this depends on an employees' self vs other orientation. Individuals who emphasize others rather than oneself place importance on fitting in with peers, and gaining a sense of belonging. Moreover, the self-ratings by other-oriented individuals not only showed greater agreement with ratings provided by their supervisor, but in addition had less leniency relative to their supervisor's appraisals of them.

These findings are consistent with previous cross-cultural studies. Markus and Kitayama (1991) found that Asians exhibit less leniency bias in their self-ratings. They also found that individuals from Eastern cultures are more likely to exhibit a greater discrepancy between their actual and ideal self. They concluded that individuals from Eastern cultures are more likely to accept information about their poor performance than employees in the West. Japanese employees, for example, are encouraged to focus on self-criticism as a way of improving their performance.

The effect of a person's self-esteem on self-ratings was examined in a laboratory study involving undergraduate students. Bernichon, Cook, and Brown (2003) found that individuals with high self-esteem tend to seek self-verifying feedback, even if it is negative. Low self-esteem individuals tend to seek positive feedback, even if it is non-self-verifying.

Ostroff et al. (2004) examined the relative importance of a number of demographic and contextual variables on self-other agreement. In a field study involving over 4000 US managers across 650 organizations, they found that women, whites, younger managers, those with less experience, and those with higher education tend to show agreement between self-ratings and ratings from

others. Over-raters tended to be men, older managers, non-whites, those with a lower education, those in service-oriented functional areas, and those who supervise a large number of employees.

The importance of agreement between one's self-appraisal and appraisals by others was shown in a study by Fletcher (1997). People whose self-appraisals are relatively aligned with the appraisals they receive from others are usually high performers. This, Fletcher concluded, is due to their 'self-awareness' regarding their ability.

The extent to which self-monitoring and attitudes affect rating accuracy was subsequently examined. Miller and Cardy (2000) found that the self-ratings of self-monitoring individuals are higher than their low self-monitoring counterparts. Furthermore, they found that there is higher agreement with other sources of appraisals for low self-monitoring than for high self-monitoring individuals. Subsequently, Jawahar (2001), in a field study involving university employees, found that rating accuracy is higher for low rather than for high self-monitors. In short, self-monitoring significantly influences rating accuracy such that one's accuracy declines with increasing levels of self-monitoring.

Mann and Budworth (2005) found that at least three variables affect the extent to which there is agreement between self- and peer appraisals of counterproductive behavior in groups of undergraduate students who worked together for 12 weeks. Consistent with Rotundo and Sackett (2002), they defined counterproductive behavior as voluntary behavior that harms the well-being of the group. They found that agreement increases to the extent that self and peers score similarly on a measure of conscientiousness, have common beliefs as to what constitutes integrity, and engage in similar levels of counterproductive behavior.

Schrader and Steiner (1996) found that an effective way to improve agreement between self- and supervisory assessments of performance is to use explicit absolute criteria rather than internal comparisons with self, or relative comparisons with others. For example, in a study involving over 3000 supervisors, 12,000 peers, and nearly 3000 subordinates, relative agreement among sources was obtained when a 48-item behaviorally oriented appraisal instrument was used (Facteau & Craig, 2001).

Multisource Feedback

Rather than choosing only one of the above sources of appraisal, 90% of Fortune 1000 firms now collect feedback from multisources (Atwater & Waldman, 1998). While some organizations use multisource feedback for developmental purposes only, others use them for making administrative decisions regarding compensation, job placement, and promotion (London & Smither, 1995). A survey of 100 organizations in Canada revealed that over half were either using a multisource system, or were considering using it in the near future (Brutus & Derayeh, 2002). Basing an appraisal of a person on information from the above

sources is referred to in the literature as 360-degree feedback because it provides a relatively complete picture of an employee's performance.

In the Psycinfo database, there are no references to 360-degree feedback prior to 1990, and only four citations between 1990 and 1995, namely four unpublished doctoral dissertations. Since 1995 there have been at least 89 studies on input from multisources, namely, supervisor, peers, subordinates, and, in many instances, self (Bracken, Timmreck, & Church, 2001). This is likely because input from multisources takes into account the 'multidimensional nature of jobs as seen by different constituencies' (London & Smither, 1995, p. 804). It takes into account the fact that different populations (e.g., peers, subordinates) have different opportunities to observe different aspects of a person's performance (Murphy & Cleveland, 1995). Similarly, Borman (1997) argued that appraisers from different populations (e.g., supervisors vs subordinates) attach different weights to the same aspects of performance that they observe as a result of their different organizational perspectives. Thus collecting appraisals from multisources provides an integrated, holistic view of an employee, thereby offsetting the drawbacks of an appraisal from only a single vantage point (e.g., supervisor). Moreover, 360-degree feedback is consistent with the organizational environment in this new millennium, one that is increasingly filled with novel, ill-defined problems. As Latham and McCauley (2005) observed, tomorrow's issues are unlikely to lend themselves to rapid solutions by assigning them to a single leader. Multiple perspectives and expertise will have to be brought to bear, especially due to the multiple stakeholders who will claim a stake in these complex issues. Hence the collective perspective of multiple individuals will often be required on performance-related decisions.

Bailey and Fletcher (2002) conducted a study of 104 managers where input was collected from at least one supervisor, three subordinates, as well as a self-assessment. There was agreement among sources on the manager's competence, developmental needs, and between the manager's self-evaluation and the evaluations the manager received from others. Agreement among multiple sources of an appraisal is evidence of convergent validity or accuracy (Bozeman, 1997).

With regard to interobserver reliability within populations of appraisers, peers have the highest, followed by subordinates and supervisors (Greguras & Robie, 1998). Thus there is benefit in treating each appraisal source as a measure of a different aspect of an employee's performance for determining developmental goals, and/or making an administrative decision (e.g., promotion).

Intersource agreement among populations, however, is not necessarily to be expected (Murphy & Cleveland, 1995). This is because each source of appraisal (e.g., peers, subordinates) often observes an employee in different contexts (Bozeman, 1997). The correlations among ratings made by a supervisor, peer, subordinate, and self range from a high of 0.79 (supervisor–peer) to a low of 0.14 (subordinate–self). Disagreement in conclusions are especially likely

if different appraisal instruments are designed to assess those aspects of the job that a specific population (e.g., subordinates) is most likely to observe the employee performing on an ongoing basis (Facteau & Craig, 2001). Thus a person's appraisal may differ across populations (supervisor/subordinates) because of actual differences in the type of behaviors that are observed. Hence, Atwater, Ostroff, Yammarino, and Fleenor (1998) argued the importance of taking into account multiple appraisal sources in assessing and explaining a person's effectiveness.

Bailey and Fletcher (2002), cited earlier, were among the first to examine how well multisource ratings predict a person's subsequent performance. They found that over a two-year period both employees' and managers' assessments of managerial competence increased while perceived developmental needs decreased. Atkins and Wood (2002) used performance in an assessment center as a proxy for performance on the job. The average of supervisor, peer, and subordinate ratings predicted performance, as did the supervisor ratings alone. With respect to self-ratings, those who rated themselves most highly tended to be the poorest performers. Conversely, those who rated themselves in the midrange of the scale were more likely to be higher performers than those who rated themselves at the top or bottom ends of the scale. In other words, highly competent employees underestimated their abilities while less competent employees tended to overestimate their abilities relative to others. Atkins and Wood urged caution in the use of self-ratings in a 360-degree feedback program. They suggested that self-ratings should not be interpreted as reflecting actual competency levels, and should not be included in aggregations of ratings for evaluation purposes.

With respect to user reactions, Maurer, Mitchell, and Barbeite (2002) found that positive attitudes toward the system is dependent upon a positive relationship between peer and subordinate ratings and peer–self agreement, such that the more different peer ratings are compared to self-ratings, the more favorable ratee attitudes are toward the system. Other predictors of positive attitudes toward a 360-degree appraisal include a work context that involves people who are supportive of skill development, and beliefs by the feedback recipients that it is not only possible for them to improve their skills, but that they are capable of doing so.

No appraisal method is above criticism. A consulting firm, Watson Wyatt (2002), reported that 360-degree feedback can lead to a 10.6% decrease in corporate shareholder value. This finding may reflect a failure by these firms to take into account individual and contextual factors that affect receptivity to 360-degree feedback. Funderburg and Levy (1997) found that organizations who have individuals with high self-esteem, internal locus of control, individuals who report to autocratic leaders, and individuals who work in an environment that encourages seeking feedback have a positive attitude toward 360-degree appraisals. The finding regarding people with autocratic leaders reacting favorably to 360-degree feedback is ironic. It may reflect the fact that

those with highly participatory leaders receive feedback on an ongoing basis and hence perceive less need for information from multisources.

In an essay, Ghorpade (2000) worried that the use of multiple sources of an appraisal may nevertheless fail to take into account situational factors that adversely affect a person's performance. Hence, more information through the use of multisources would not necessarily yield superior feedback. Empirical research suggests otherwise.

Church (2000) found that managers who received more favorable multi-source feedback had lower employee turnover and higher service quality in their workgroups. In addition, Church found that multisource ratings (e.g., supervisors, peers, and subordinates) accurately distinguished between high- and low-performing managers as assessed by independent measures. Erikson and Allen (2003) found that multisource feedback ratings were related positively with retail store revenue, gross margin, and sales of accessories.

Smither, London, and Reilly (2005) presented a theoretical framework and reviewed empirical evidence suggesting that an improvement in performance resulting from multisource feedback is moderated by a number of variables. These variables include a person's perception that there is a need to alter one's behavior, belief that the change is feasible, and goal setting based on this feedback.

In summary the literature suggests that 360-degree feedback should be used if: (a) the appraisal instrument is based on the core behaviors directly relevant to the organization; (b) peer and subordinate appraisals are anonymous; (c) the organizational environment encourages seeking feedback; and (d) appraisers are trained in conducting appraisals. The latter topic is discussed next.

TRAINING APPRAISERS

The importance of training appraisers is further evident in a study of auto-mobile salespeople. Pettijohn, Pettijohn, and d'Amico (2001) reported that job satisfaction was due in part to appraisal-related factors. Specifically, they found that—to the extent that there are clear criteria, the criteria meet the employee's approval, and the resulting appraisal of the person on these criteria is perceived as fair—job satisfaction increases. This finding is supported by another study that shows that a poorly implemented performance appraisal system increases employee 'burnout' (Gabris & Ihrke, 2001). A study involving the Australian government showed that important predictors of employee acceptance or rejection of the appraisal program is satisfaction with their supervisor, the feedback anticipated by the employee, and anticipated consequences of the appraisal (Langan-Fox, Waycott, Morizzi, & McDonald, 1998).

Further evidence of the need for training on ways to increase objectivity and accuracy in making assessments of people is the finding that personality differences among appraisers can affect an employee's appraisal. Kane, Bernardin,

Villanova, and Peyrefitte (1995) found that leniency in ratings is a relatively stable characteristic of a rater. Bernardin, Cooke, and Villanova (2000), in a laboratory study involving undergraduate students, found that two of the Big Five personality factors predict leniency ratings. People who scored highly on agreeableness tended to provide more elevated ratings of a peer's performance on group exercises, whereas individuals who scored highly on conscientiousness rated peers relatively low.

Even an appraiser's mood can affect rating accuracy. Freid, Levi, Ben-David, Tiegs, and Avital (2000), in a field study conducted in the USA and Israel, found that negative mood predisposition was negatively associated with performance ratings of employees whereas a positive mood predisposition was unrelated to performance ratings of employees.

Not surprisingly, organizational politics affect appraisals. Tziner (1999), in a field study of police officers, found a negative relationship between an appraiser's self-efficacy as a rater and the effect of political considerations. People who perceive themselves as insufficiently skilled as raters tend to remove themselves from the uncomfortable situation of appraising others by invoking political considerations and distorting the results of the performance appraisal. This reflects the raters' desire to avoid giving performance ratings that might antagonize employees or provoke their resentment. Tziner, Latham, Price, and Haccoun (1996) have developed a 25-item scoring instrument to measure perceptions of the extent to which performance appraisals are affected by organizational politics.

Uggerslev and Sulsky (2002) found that supervisors take into account information from an employee's coworkers when making an appraisal. But, they do so only when it is consistent with their own observations, and they do so primarily when the information is positive.

A final reason for training appraisers can be found in a field study conducted in Canada, Israel, and the USA. Tziner, Murphy, and Cleveland (2001) found that an appraiser's self-efficacy for making accurate ratings as well as confidence, trust, and comfort level with the appraisal process accounted for significant variance in raters' tendency to differentiate among employees and to differentiate their ratings in accordance with the different dimensions of performance on which an employee was being assessed.

Rater Accuracy

Training that allows people to evaluate actors presented on videotape receive feedback as to their rating accuracy, and to practice, practice, practice increases objectivity and minimizes leniency, halo, and similar to me biases (Cardy & Keefe, 1994; Latham, Wexley, & Pursell, 1975; Noonan & Sulsky, 2001; Schleicher & Day, 1998; Sulsky & Keown, 1998; Woehr & Huffcutt, 1994). This training teaches appraisers (a) the relevant performance criteria for evaluating people, (b) the relevant job behaviors, and (c) ways to effectively

minimize errors in judgment when using the appraisal instrument. In addition, Keown-Gerrard and Sulsky (2001) found that this training is highly effective when it includes context-specific issues relevant to the organization and/or job, particularly situational constraints that inhibit or interfere with a person's performance. Further research in this area is not likely to be needed. Findings from this research can be put into practice.

DeNisi and Peters (1996) conducted the only field study of cognitive processes on performance appraisals. They found that structured diary keeping and structured recall results in raters having more positive reactions to the appraisal process, having better recall of performance information, and producing ratings that are less elevated. Structured recall involves asking raters to recall performance incidents and then to record them according to a given format. Organizing information according to persons, rather than according to task dimensions, was better for both ratings and recall. Organizing information according to persons involves recalling incidents for one subordinate and then moving on to the next employee. Raters using this format reported that their ratings were more fair and accurate. As Arvey and Murphy (1998) noted, no direct measurement of the proposed cognitive processes was assessed in this study, but the implications for practice are straightforward.

Similarly, Jelley and Goffin (2001) found that priming a rater improves the accuracy of performance ratings. The experiment involved undergraduate students making a global performance rating of their lecturers on dimensions such as clarity and enthusiasm. This was followed by a request for them to assess these same lecturers on the specific behaviors that define these two performance criteria. Replications in the field are now needed.

Tests of Dweck's (1999) implicit person theory (IPT) have been restricted largely to children. Her findings show that entity theorists believe that personal attributes of an individual are largely fixed. This leads them to quickly form strong impressions of others that they resist revising, even in light of contradictory information. Incrementalists, on the other hand, appreciate the dynamic and personal determinants of behavior, and thus reconsider their initial impressions after receiving new information. Based on Dweck's findings, Heslin, Latham, and VandeWalle (2005) examined the extent to which the IPT of managers affects their appraisals of others. They found that the extent to which managers hold an incremental IPT is positively related to their recognition of both good and poor performance, relative to the employee's behavior they initially observed. The judgments of managers with entity theorist beliefs are anchored by their prior impressions of an employee.

The authors then adapted Aronson's (e.g., Pratkanis & Aronson, 2001) self-persuasion techniques to train managers with entity beliefs to adopt an incremental IPT. The long-term benefit of this training was shown in a six-week follow-up study. Those entity theorists who were randomly assigned to the training condition acknowledged an actual improvement in an employee's performance relative to their counterparts in the control group.

PERFORMANCE MANAGEMENT

The present review shows that significant advances have been made by behavioral scientists in the past century regarding the development and implementation of performance appraisal systems. A primary objective of doing an appraisal is to inculcate in employees the desire for continuous improvement. The attainment of this objective, however, is hampered by the fact that appraisals are done on a discrete basis. That is, they are usually done quarterly, bi-annually, or annually. Currently, there is a major shift in emphasis from performance appraisal to performance management, from being a performance appraiser to becoming a performance coach (Cederblom & Pemerl, 2002). The shift is from a discrete activity to one that is an ongoing process. The shift is from being solely an evaluator of people to, in addition, being a developer of people.

In a study of 400 companies, Campbell, Garfinkel, and Moses (1996) found that cyclical year-round performance management (i.e., feedback, analyzing results, goals) effectively increases organizational performance. In a study of executive coaching in a public sector municipal agency, Olivero, Bane, and Kopelman (1997) found that those who received coaching that included goal setting, collaborative problem solving, practice, feedback, and evaluation of end results increased their productivity dramatically. Consistent with these findings, Smither et al. (2003), in a field study of over a thousand senior managers, found that those who worked with an executive coach, in an organization that uses a multisource feedback system, were more likely than other managers to set specific high goals, and to seek feedback on ways to improve their performance from their supervisors. This led to an improvement in the ratings they received from their direct reports as well as from their supervisor.

Luthans and Peterson (2003), in a field study within a small manufacturing company, found that a way to improve the effectiveness of a 360-degree system is to combine it with coaching that focuses on enhancing self-awareness and performance management. The combination of feedback and coaching resulted in improved manager and employee satisfaction, increased commitment to the organization, reduced intentions to quit, and indirectly increased the firm's performance.

In short, coaching appears to be a powerful catalyst for bringing about a relatively permanent improvement in a person's behavior (Kalinauckas & King, 1995), because a coach challenges employees on a day-to-day basis, to instill in them the confidence that they can expand their abilities to attain desired goals (Ellinger, Watkins, & Bostrom, 1999). Hence, communication skills are a core competency of an effective coach (Richardson, 1998) especially with regard to conveying knowledge as well as expectations to others (Tyler, 1997). The question that has yet to be answered is, who is an ideal coach? Is it the person's supervisor, peers, an external agent, or oneself?

Hall, Otazo, and Hollenbeck (1999) defined a coach in an organizational setting as a person who works with others to develop and implement strategies

to improve their performance. Typically, the coach is not the employee's supervisor and hence does not provide, and is not solicited for, input regarding the organization's formal reward system for the employee. For example, a global consulting firm, with offices in such countries as Australia, Canada, and the United Kingdom, uses self as well as peer coaching due to the fact that it is a partnership, and the partners work alone as well as in teams on assignments. The emphasis is on increasing the person's interpersonal skills within the firm as well as with the client in order to increase the firm's revenue. A nuclear power plant in Canada uses an external agent to coach its key managers on ways to increase their interpersonal skills, especially team playing, as does an investment bank in the USA.

Despite the voluminous practitioner literature on the value of coaching, there are few or no empirical studies that have evaluated the effectiveness of this practice on the subsequent performance of others. Moreover, there is little or no empirical evidence as to who is most effective as a coach. Theoretical and empirically derived frameworks, however, do exist for guiding research on this topic. For example, evidence suggestive of the effectiveness of self-coaching can be inferred from Aronson's (1999a, 1999b) self-persuasion theory. The theory states that self-persuasion strategies produce more powerful and long-lasting effects than do alternative sources. Attitude and behavior change induced from others is relatively short term, especially when there is a strong emotional component (e.g., an adverse effect on one's job or career). With self-persuasion, there is no direct attempt from others to convince anyone to do anything. Hence, the theory states that self-persuasion allows individuals to convince themselves of the desirability of a behavior or behaviors. The empirical data in support of this theory, however, have been limited to social psychology experiments involving such diverse areas as use of condoms by teenagers (Aronson, Fried, & Stone, 1991; Stone, Aronson, Crain, Winslow, & Fried, 1994), and energy and water conservation among adults (Dickerson, Thibodeau, Aronson, & Miller, 1992). With the exception of the study by Heslin et al. (2005) described earlier, there are few or no studies of the effect of self-persuasion in organizational settings.

Further evidence suggesting the efficacy of self-coaching as a self-regulatory technique comes from Meichenbaum (1977), a clinical psychologist who developed a method for changing a person's dysfunctional to functional self-talk. Millman and Latham (2001) effectively adapted this methodology for training displaced managers to use this verbal self-guidance (VSG) technique to successfully obtain re-employment. In two subsequent studies, Brown and Latham (in press) found that VSG increased a person's team-playing behavior within a group, and Brown (2003) found that it increases both collective efficacy as well as a team's performance.

Evidence suggesting the effectiveness of peers as coaches can be inferred from Festinger's (1954) social comparison theory. The theory states that the drive for self-comparison is a force acting on a person to belong to a group. On

subjective criteria (e.g., team playing), people assess their ability in comparison to others. 'Given a range of possible persons for comparison, someone close to one's ability or opinion will be chosen for comparison . . . Those with whom one does not compare oneself are different kinds of people or members of different groups' (Festinger, 1954, p. 121).

Support for the use of peers as coaches can also be found in sociotechnical systems theory (Trist, 1977). The theory states that it is the group who should monitor the individual's contribution. Productivity is fostered by the group allocating tasks and other rewards and punishments to control what the group considers to be a fair contribution by a group member (Emery & Thorsrud, 1976). Little, if anything, however, has been published in the sociotechnical literature on the effectiveness of peers in increasing the performance of a colleague.

Support for the use of an external agent as a coach can be inferred from the social psychology literature on persuasion. For example, Cialdini (2001) argued that authority is a key determinant of another person's attitudes and actions. He argued further that a person can harness the power of authority by touting experience, expertise, and credentials. People value the expertise of authorities because it helps them to choose both quickly and well. Expertise refers to the extent to which a person is perceived to be a source of valid assertions, especially with regard to the task that is being performed. The higher the perceived source credibility, the higher the likelihood that behavior will change as a result of it (Hovland & Weiss, 1951). There is a positive relationship between the credibility of the source of delivery and information retention (Zagona & Harter, 1966), feedback acceptance (Halperin, Snyder, Shenkel, & Houston, 1976), feedback favorability (Albright & Levy, 1995), intention to use feedback (Bannister, 1986), and performance in a laboratory setting (Northcraft & Earley, 1989). Thus on the basis of the social psychology literature, one can infer that the use of an external agent as a coach is effective in bringing about a desired behavior change in others.

An initial step to providing an answer to the above question was taken by Sue-Chan and Latham (2004). Two studies in two different continents using two different dependent variables examined the relative effectiveness of external, peer, and self-coaches on the performance of participants in two MBA programs. The first study involved MBA students in Canada. Those who were coached by an external coach exhibited higher team-playing behavior than did those who were coached by peers. The second study involved EMBA managers in Australia. Those who were coached by an external person or who were self-coached had significantly higher grades than those who were coached by a peer. In both studies, an external coach was perceived to have higher credibility than peers. In the second study, self-coaching was perceived to be more credible than coaching from peers. Satisfaction with the coaching process was highest among the managers who had an external coach.

Future research should examine the effect of peers serving as coaches in work contexts where the peer has as much or more expertise than the person who is

being coached. In such a context, a peer who is viewed as an expert is likely to be a much more credible coach than was the case in the two present studies. That peers in the two present studies were not effective coaches appears to have been due to their perceived lack of expertise.

Participants in the self-coaching condition in the first study stated that they had difficulty identifying which behaviors to improve on in the future. The Australians who coached themselves, however, reported that the process was effective. As one Australian manager noted, 'self-coaching raised my awareness of "positive" and "bad" behaviors.' Thus, the contrary findings of the present two studies suggest a boundary condition for Aronson's (1999a, 1999b) theory of self-persuasion, namely, the extent to which the person possesses the knowledge and ability to perform the task. In all of Aronson's studies of self-persuasion, the person already possessed the requisite knowledge and skill to make the behavior change. The issue confronting the individual was primarily one of motivation to do so. The students in Canada lacked the knowledge and skill necessary to improve themselves as team players. The Australians were experienced managers from industry. They had little difficulty in coaching themselves using the behavioral appraisal instrument provided to them for attaining a high grade. Studies are now needed to compare the self-coaching effects of VSG training with Aronson's self-persuasion techniques alone and in combination.

The movement away from performance appraisal to performance management necessitates managers enlarging their role to include the coaching of employees. In a longitudinal field study, Heslin, VandeWalle, and Latham (2005) found that a manager's IPT predicts employee ratings of their coaching behavior. That is, managers who hold incremental beliefs coached their subordinates. A second study showed that managers with entity beliefs who were trained to adopt incremental beliefs provided more improvement suggestions of higher quality than their colleagues in the control group. It would appear that IPT is a state that is malleable rather than relatively fixed.

Feedback

Regardless of who serves in the role of coach, feedback is central to the effectiveness of the performance management process. Yet a meta-analysis by Kluger and DeNisi (1996) on the effect of feedback showed that in one-third of the interventions, feedback decreased rather than increased performance. However, when feedback is delivered in an informational rather than in a controlling way, a person's subsequent performance improves (Zhou, 1998). In addition, for feedback to bring about a positive change in behavior, it must: (a) focus on the behavior rather than the person; (b) be selective so as not to overwhelm the person; (c) focus on the behavior that is desired, and the way to demonstrate it; and (d) be the basis for setting specific high goals (DeNisi & Kluger, 2000).

Nease, Mudgett, and Quinones (1999) found that people who have high self-efficacy on the requisite tasks disregard repeated negative feedback. Not surprisingly, those with low self-efficacy did not reject feedback that was repeatedly negative.

Brett and Atwater (2001), in a study involving graduate students, found that negative feedback, which was lower than the person expected, was not viewed as accurate or useful. Not surprisingly, it led to feelings of anger and discouragement. High ratings, however, were not related to positive reactions, but only to the absence of negative reactions. These feelings suggest that people who may need feedback the most because they are not performing well are not receptive to it.

Lam, Yik, and Schaubroeck (2002) examined the effects of performance appraisal feedback on job and organizational attitudes of tellers in a large international bank. An employee's negative affectivity moderated the relationship between the receipt of favorable feedback and subsequent job attitudes. That is, there was an improvement in an employee's attitude following favorable feedback that persisted for six months among tellers who scored low on a measure of negative affectivity. This did not occur for those who were high on negative affectivity.

Smither and Walker (2004), in a study of 176 managers, found that those who received favorable comments tended to improve more than managers who received unfavorable comments. Contrary to expectations, behavior/task-oriented comments did not lead to greater improvement than did trait-focused comments. In addition, those who received fewer unfavorable comments improved more than did other managers. With respect to the specificity of the feedback, Goodman, Wood, and Hendrickx (2004) found that the more systematic a person is in the exploration process, the less confounded the information obtained, and the more beneficial the feedback for performance.

Frese (2005) has shown that people can be easily taught through instructions to embrace negative feedback by framing errors as beneficial to the learning process, and to be resilient subsequent to making an error, through systematic exploration. Future research should examine the effect of setting specific high-learning goals in the context of error management training.

The impersonal nature of communicating via e-mail appears to increase the objectivity of an appraisal, but it lessens the sensitivity and tact of the appraiser (Weisband & Atwater, 1999). Sensitivity and tact in giving feedback is critical for bringing about a change in behavior regardless of the medium by which it is communicated. For example, Seifert, Yukl, and McDonald (2003) found that having a competent, supportive facilitator present when feedback is provided, vs receiving only a printed feedback report, increases the perceived utility of the feedback, and results in more behavior change.

Employees should be encouraged to seek feedback. Seeking feedback has been found to increase an employee's understanding of where performance improvement is needed, and hence it increases goal commitment (London, Smither, & Adsit, 1997; Walker & Smither, 1999). Moreover, appraisers engage

in fairer behavior when interacting with and justifying an appraisal decision with an assertive rather than an unassertive employee (Korsgaard, Roberson, & Rymph, 1998).

Goal Setting

Feedback in the absence of goal setting has little or no effect on behavior (Locke & Latham, 2002), because feedback is only information; its effect on action depends on how it is appraised by the recipient, and what decisions are made with respect to it. For feedback to improve behavior, specific high goals must be set (Latham, Locke, & Fassina, 2002), because goal setting affects choice, effort, and persistence. A meta-analysis revealed a strong relationship between goal commitment and a person's performance (Klein, Wesson, Hollenbeck, & Alge, 1999).

The importance of goal commitment was also shown in a field study by Renn (2003) involving rehabilitation counselors of a state agency. Feedback had a positive relationship with work performance only for those individuals with high goal commitment; it had a negative relationship with performance for those with lower goal commitment.

A key variable for bringing about goal commitment is to focus on outcome expectancies (Klein et al., 1999; Latham, 2001). That is, getting people to see the relationship between what they do or fail to do and the outcome they can expect. Employees who do not see the link between improving their performance and an improvement in their appraisal have little or no motivation to change their behavior (Boswell & Boudreau, 2002). This finding provides additional evidence for the necessity of training appraisers on ways to improve the accuracy of their evaluations.

The importance of goal-setting theory was explained by Locke and Latham (1990). Relative to exhorting people to 'do their best,' the assignment of a specific high goal increases performance by 0.42 to 0.80 standard deviations. A meta-analysis by O'Leary-Kelly, Martocchio, and Frink (1994) found an average standard deviation difference of 0.91 in the performance of teams *with* vs *without* a specific high goal.

Recent research on goal setting shows that when the person has the requisite knowledge and skill, specific high outcome goals should be set; when the person lacks the requisite knowledge and skill, specific high-learning goals should be set (Seijts & Latham, 2001, 2005; Winters & Latham, 1996). For example, a golfer with a low handicap might set an outcome goal of 68. A golfer with a high handicap might set a specific high goal of discovering (learning) six ways to lower the handicap. The focus of learning goals is on the process or procedure for attaining the outcome rather than on the outcome itself. These findings have yet to be studied in the context of performance appraisals.

Behaviorally set goals should be set when they are based on an appraisal instrument derived from a job analysis. In such cases, the behavioral items on the

appraisal instrument removes the necessity of discovering the requisite knowledge or behavior for performing tasks effectively (Brown & Latham, 2002).

When the environment is inherently unstable, proximal or sub-goals should be set in addition to a distal goal (Latham & Seijts, 1999) because, in highly dynamic situations, it is important to actively search for feedback and react quickly to it (Frese & Zapf, 1994). In addition, Dorner (1991) has found that performance errors on a dynamic task are often due to deficient decomposition of a goal into proximal goals. Proximal goals can increase what Frese and Zapf (1994) call error management. Errors provide information to employees as to whether their picture of reality is congruent with goal attainment. There is an increase in informative feedback when proximal or sub-goals are set relative to setting a distal goal only.

In addition to being informative, the setting of proximal goals can also be motivational relative to a distal goal that is far into the future. The attainment of proximal goals increases commitment, through enactive mastery, to attain the distal goal (Bandura, 1997). These findings regarding learning as well as proximal goals have yet to be examined within the context of performance appraisals.

Tziner, Joannis, and Murphy (2000), however, compared three appraisal instruments to determine the extent to which they facilitate goal setting. Goals set with BOS were more specific than those established on the basis of BARS or trait scales.

Organizational Justice

Regardless of how well feedback is delivered, and how goals based on this feedback are set, few things demoralize a person or team faster than feelings of jealousy, perceptions of favoritism, or beliefs that someone is unfairly getting a 'better deal' than someone else. Organizational justice theory states that in addition to being fair, decision makers must be seen as being fair (Greenberg, 2000). Providing the logic or rationale for decisions increases perceptions of fairness.

Evidence presented as to the validity of appraisal decisions increases perceptions of procedural justice; providing feedback that has taken into account the employee's input increases perceptions of interactional justice as well as the belief that one's voice has been heard (Erdogan, Kraimer, & Liden, 2001). As noted earlier, a meta-analysis involving over 6000 employees found that allowing employees a voice in the appraisal process increases employee satisfaction with the appraisal, perceived fairness of the appraisal, perceived usefulness of the appraisal, and the employee's motivation to improve as a result of the appraisal (Cawley et al., 1998).

Taylor, Tracy, Renard, Harrison, and Carroll (1995) examined the differences in outcomes resulting from a 'traditional' appraisal vs one that is based on 'due process.' The latter is defined as giving people adequate notice of their appraisal (e.g., explaining the standards in advance, seeking self-appraisals, giving

feedback on an ongoing basis), fair hearing (adequate observations of the person's performance, granting an opportunity to explain self-evaluations), and judgment based on evidence (e.g., consistent application of standards, opportunity to appeal). Employees who were appraised in the due process condition perceived their appraisal system to be more fair and accurate than those in the traditional condition. This occurred despite the fact that employees in the due process condition had lower evaluations than those who received a traditional appraisal. This suggests that training appraisers in the principles of organizational justice is also an effective way of minimizing leniency error. Levy and Williams (1998) found that perceived system knowledge—that is, the extent to which individuals perceive that they understand the objectives of the appraisal system, how the appraisal process works, and the goals of the process—predicts user reactions, perceptions of fairness, job satisfaction, and organizational commitment. Lam and Schaubroeck (1999), in a field study of front-line supervisors in Hong Kong, compared outcome-focused appraisals with process-focused appraisals, that is, one that provides information on the manner in which an individual implements the organization's strategy. Process-focused appraisals had a more positive impact on satisfaction with the appraisal, perceived accuracy of the appraisal, and expectations of performance improvement.

MOTIVATING APPRAISERS

For appraisers to be motivated to take appraisals seriously, they must see the relationship between doing so and desirable outcomes for themselves, and the organization (Latham & Latham, 2000). They must see the relationship between the performance appraisal and other human resource systems including staffing, training, and, as emphasized throughout this chapter, the effective implementation of the organization's strategic plan. Additional research is needed on outcome expectancies. A maxim attributed to the late Mason Hare is 'that which gets measured gets done.' Hence many organizations have embraced a balanced scorecard (Kaplan & Norton, 1993**).** PriceWaterhouseCoopers, for example, places equal weighting on client (revenue generated from new and existing clients; client satisfaction), people (coaching and developing staff), and firm (service on committees crucial to the ongoing reinvention of the firm's business strategy necessary to ensure its success in a local and global marketplace).

NEW FRONTIERS

What are the next steps for making significant advances in the effectiveness of performance management? The answers may lie in the fields of decision making, motivation, and clinical counseling psychology, because there are at least

two interrelated reasons for conducting performance appraisals of people—namely, to make developmental decisions regarding the ongoing improvement of an employee's performance, and to make administrative decisions regarding retaining, promoting, transferring, paying, and/or terminating the person based on the extent to which the person's performance is contributing to an organization's strategy. These are unarguably important decisions, yet the knowledge accumulated from theory and empirical research in the area of decision making has been largely ignored in the context of performance appraisals. There is at least one study that suggests that it is a mistake to continue doing so.

Latham and Budworth (2005) found that managers make an anchoring error in their appraisals of others. In a laboratory experiment, those who were randomly assigned to a condition where they were told how good they were subsequently rated a hypothetical employee shown on a videotape significantly higher than those managers who were told that they were poor performers. A field study, involving a manufacturing company, showed that there is a significant correlation between the performance rating that a manager receives and the rating the manager subsequently gives to his or her direct reports. Further research would appear necessary on the dangers of anchoring and other decision-making errors (e.g., satisficing, saliency, escalation to commitment, hindsight bias) as well as on ways to overcome them.

Motivation

Performance management involves coaching people in ways that will motivate them to improve their performance. Latham and Pinder (2005) reviewed the literature on motivation in the workplace. The study of traits has recaptured the attention of researchers. For example, Baum, Locke, and Smith (2001) showed that the relationship between personality and performance is mediated by situationally specific goals and self-efficacy. Ways of setting goals and building self-efficacy should become the mainstay of the coaching process. The former is a strong variable that masks individual differences (Locke & Latham, 2002).

In a study of more than 19,000 participants in 25 countries, Scholz, Dona, Sud, and Schwarzer (2002) found a high degree of consistency in the psychometric properties of measures of self-efficacy, a variable key to goal commitment and self-regulation. Steers and Sanchez-Runde (2002) developed a conceptual model that has yet to be empirically tested regarding leading across diverse cultures. The model focuses on self-efficacy beliefs, work motivation levels and goals, as well as the nature of incentives and disincentives to perform well, and the extent to which societal culture is a moderator.

Seijts, Latham, Tasa, and Latham (2004) discovered that goal orientation is more properly treated as a state rather than a stable individual difference variable. Hence, coaches can inculcate a learning goal orientation in employees in addition to setting specific high-learning goals. A promotion focus (Higgins, 1999) is conceptually similar to a learning goal orientation. Erez, Kaplan, and Van-Dijk (2004) found that it too is a malleable state rather than a stable trait.

Research is needed on whether there are incremental increases in an employee's performance when a specific high-learning goal is set, and a learning goal orientation as well as a promotion focus is conveyed by one's coach.

Significant findings in social psychology regarding implementation intentions and automotive goals should be investigated within the context of performance management. Social psychologists are currently examining the variety of ways in which goals may be automatically activated in everyday settings, and how goal priming may affect the goals a person pursues. Specifically, they are examining aspects of the environment that might cause a person to spontaneously rather than purposefully pursue a goal. In addition, they are examining how a person's self-regulation system deals with the variety of goals that may be primed by one's environment.

Bargh and his colleagues primed goals by taking advantage of their associations to semantically related words. Shah (2005) examined two other associations: (1) goals associated with those things that help to bring about their attainment, that is, instrumental goal priming; and (2) goal association with other individuals, that is, interpersonal goal priming.

A particular goal is often pursued in particular settings, and while one is engaging in specific behaviors or activities, the goal may become associated with such settings. The instrumental or means–goal association is not dependent on the semantic relation of the means and goal, but rather on their perceived functional relationship—that is, said Shah, the degree to which the means is seen as facilitating the attainment of one's goal. The stronger the association, the higher the likelihood that encountering the setting, or the means to attain it, will automatically lead to the pursuit of that goal. In short, Shah has argued that surrounding a person with various means to attain a specific goal is likely to move that goal to the center of a person's attention, thus enhancing that individual's commitment to attaining it.

In addition to semantic and instrumental goal priming, a third source of goal priming is the people who are significant to an individual. Shah (2003a) found that when people are presented the name of a close significant other subliminally, that is, a person who would want them to do well on the task, the salience of the task goal increases, as does persistence and actual performance. Shah also found that priming a significant other decreases pursuit of the goal if the significant other does not support the goal, or is strongly associated with the pursuit of a different or unrelated goal.

In still another study, Shah (2003b) found that a significant other affects how a goal is consciously perceived and expressed by the person who is pursuing it. A significant other 'automatically' affects the perceived difficulty and value of a goal, and how the person feels emotionally about success and failure regarding goal attainment.

Interpersonal goal priming even affects subsequent social interactions. Fitzsimons and Shah (2005) subliminally presented words that were designed to prime a goal to achieve academically. People, randomly assigned to an experimental group, relative to their counterparts in the control group, reported

feeling significantly closer to those specific significant others who could help them to attain the goal. In addition, they indicated that they intended to spend less time with those significant others who they perceived as not facilitating academic success. In short, the environment contains inanimate signs, signals, or cues that can 'prime' or trigger the automatic pursuit of our goals. This is both an exciting and frightening finding. How many of us will want our bosses to master the implementation of these techniques?

Clinical/Counseling Psychology

I/O psychology has a long history of successfully adapting methodology from experimental as well as social psychology for studying and resolving issues in the workplace. Latham and Heslin (2003) have stressed the potential pay-offs for adapting theoretical frameworks and methods from clinical/counseling psychology. The payoffs would appear to be particularly large for increasing the effectiveness of performance management. For example, Meichenbaum's (1977) technique for teaching people to become aware of their internal dialogue and its effect on their effort and behavior was discussed earlier (Brown, 2003; Brown & Latham, in press; Millman & Latham, 2001). VSG should be investigated as a means for ongoing coaching of self.

Ellis's (1999) rational emotive therapy (RET) has been shown in clinical settings to be effective in overcoming problems due to anxiety, anger, and procrastination. RET teaches people to set realistic personal goals, and in the face of failures to redirect their cognitive and emotional energy toward developing new strategies for attaining them. RET is also effective for overcoming a desire for perfection. This methodology should be easily adaptable to performance management in the workplace.

Seligman's (1998) training technique, based on attribution theory for overcoming learned helplessness and instilling optimism in the face of repeated failures, also warrants investigation in the context of performance management. His techniques for effectively disputing one's dysfunctional beliefs include: (a) examining and questioning the evidence for their validity; (b) focusing on changeable, specific, and non-personal causes of the negative actions one experiences, rather than making attributions to permanent, pervasive, and personal causes; (c) avoiding catastrophizing the implications of one's negative attributes; and (d) distracting oneself from recurring self-defeating thoughts. This methodology is also applicable to the workplace.

Bordin's (1994) working alliance focuses on straightforward ways of enhancing the relationship between the therapist and the client. The emphasis is on steps for ensuring a mutual understanding and agreement between the two parties regarding the goal or goals of therapy, and the tasks to be pursued to attain the goal(s). These principles would appear to be highly applicable for a supervisor, team leader, or external coach working with people in organizational settings.

An inherent danger in coaching that is seldom, if ever, discussed, let alone studied in the I/O or human resource management (HRM) literature, is the dysfunctional attachment that can occur between an employee and the coach or supervisor. Bowlby's (1979) attachment theory explains how different attachment styles systematically influence how people seek and process feedback, interact with and evaluate others, and regulate their emotions. The theory suggests that an avoidant or anxious–ambivalent attachment style on the part of a supervisor or coach may result in that person encoding less information about an employee (Fraley, Garner, & Shaver, 2000) and being less able to describe an employee's behavior objectively than those with a secure attachment style. His findings have yet to be applied to the field of performance management, but nevertheless have strong implications for who should and should not take on a coaching role.

CONCLUSIONS

Science has informed practice.

1. We know how to conduct a job analysis to develop an appraisal instrument that is reliable, valid, and legally defensible. We know how to develop appraisal instruments that facilitate goal setting.
2. The reliability and validity of different sources of appraisal are known. There is growing acceptance in organizations of the importance of using multisources for an appraisal.
3. Ways of gaining user acceptance of the appraisal system/process have been identified.
4. We know how to train appraisers so that their objectivity/accuracy increases. We also know that political factors may override the goal for accuracy.
5. We know that performance management, defined as year-round coaching of an employee, is more likely to inculcate a desire for continuous improvement, and to bring about and sustain a significant increase in an employee's behavior than an annual or even a quarterly appraisal.
6. There are three theoretical frameworks to draw upon in the practice of performance appraisal/management: goal setting with regard to the importance of setting specific high-performance, learning, or behavioral goals; social cognitive with regard to increasing self-efficacy; and organizational justice with regard to minimizing findings of favoritism, exploitation, or injustice.

Practice is informing science on areas requiring research:

1. Appraiser bias remains an issue. It is likely to become a bigger issue as global companies bring together employees with different ethnic backgrounds who live in countries with different political ideologies.
2. The effect of cultural context on effective ways of appraising and coaching employees needs to be investigated.

3. Additional information is needed on who is the ideal coach. Is it someone external to the organization, the supervisor, peers, or oneself? The answers may vary as a function of context, including societal culture. The answer may be multiple sources.

4. The study of proximal vs distal goals, and learning vs behavioral vs performance outcome goals, should be studied within the context of performance appraisal/management.

5. Adapting theories and research findings from the fields of decision making and motivation as well as clinical-counseling psychology should accelerate discoveries of ways of improving the performance management process in organizational settings.

6. Technology needs to be studied as a moderator with regard to all of the above. Large, geographically dispersed groups of people, connected only by their use of communication technology such as mobile phones, text messaging, two-way pagers, and e-mail, can already be drawn together at a moment's notice, like schools of fish, to perform some collective action. In virtual teams spanning global organizations, how will this technology facilitate or inhibit performance appraisal/management?

ACKNOWLEDGMENTS

REFERENCES

Albright, M.D. & Levy, P.E. (1995). The effects of source credibility and performance rating discrepancy on reactions to multiple raters. *Journal of Applied Social Psychology*, **25**, 577–600.

Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology*, **47**, 349–356.

Aronson, E. (1999a). *The social animal* (8th edn). New York: Freeman.

Aronson, E. (1999b). The power of self-persuasion. *American Psychologist*, **54**, 875–884.

Aronson, E., Fried, C., & Stone, J. (1991). Overcoming denial and increasing the intention to use condoms through the induction of hypocrisy. *American Journal of Public Health*, **81**, 1636–1638.

Arvey, R.D. & Murphy, K.R. (1998). Performance evaluation in work settings. *Annual Review of Psychology*, **49**, 141–168.

Atkins, P.W.B. & Wood, R.E. (2002). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology*, **55**, 871–904.

Atwater, L., Ostroff, C., Yammarino, F.J., & Fleenor, J.W. (1998). Self–other agreement: Does it really matter? *Personnel Psychology*, **51**, 577–598.

Atwater, L., Roush, P., & Fischthal, A. (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology*, **48**, 35–59.

Atwater, L.E. & Waldman, D.A. (1998, May). Accountability in 360 degree feedback. *HR Magazine*, 96–104.

Bailey, C. & Fletcher, C. (2002). The impact of multiple source feedback on management development: Findings from a longitudinal study. *Journal of Organizational Behavior*, **23**, 853–867.

Bandura, A. (1997). *Self-efficacy*. New York: Freeman.

Bannister, B.D. (1986). Performance outcome feedback and attributional feedback: Interactive effects on recipient responses. *Journal of Applied Psychology*, **71**, 203–210.

Bartol, K.M. (1999). Gender influences on performance evaluations. In G.N. Powell (Ed.), *Handbook of gender and work*. Thousand Oaks, CA: Sage.

Bartol, K.M., Durham, C.C., & Poon, J.M.L. (2001). Influence of performance evaluation rating segmentation on motivation and fairness perceptions. *Journal of Applied Psychology*, **86**, 1106–1119.

Baum, J.R., Locke E.A., & Smith, K.G. (2001). A multidimensional model of venture growth. *Academy of Management Journal*, **44**, 292–303.

Beehr, T.A., Ivanitskaya, L., Hansen, C.P., Erofeev, D., & Gudanowski, D.M. (2001). Evaluation of 360 degree feedback ratings: Relationships with each other and with performance and selection predictors. *Journal of Organizational Behavior*, **22**, 775–788.

Bennington, L. & Roberts-Calvert, B. (1998). Antidiscrimination legislation and HRM practice. In M. Patrickson & L. Hartmann (Eds), *Managing an ageing workforce*. Sydney, Australia: Woodslane Publishing.

Bennington, L. & Wein, R. (2000). Anti-discrimination legislation in Australia: Fair, effective, efficient or irrelevant? *International Journal of Manpower*, **21**, 21–33.

Bernardin, H.J. (2005). Behavioral observation scales. In S. Cartwirght (Ed.), *Human resource management. V: The Blackwell Encyclopedia of Management* (2nd edn). Malden, MA: Blackwell.

Bernardin, H.J., Cooke, D.K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology*, **85**, 232–234.

Bernichon, T., Cook, K.E., & Brown, J.D. (2003). Seeking self-evaluative feedback: The interactive role of global self-esteem and specific self-views. *Journal of Personality and Social Psychology*, **84**, 194–204.

Bettenhausen, K.L. & Fedor, D.B. (1997). Peer and upward appraisals: A comparison of their benefits and problems. *Group and Organization Management*, **22**, 236–263.

Biernat, M., Crandall, C.S., Young, L.V., Kobrynowicz, D., & Halpin, S.M. (1998). All that you can be: Stereotyping of self and others in a military context. *Journal of Personality and Social Psychology*, **75**, 301–317.

Boldry, J., Wood, W., & Kashy, D.A. (2001). Gender stereotypes and the evaluation of men and women in military training. *Journal of Social Issues*, **57** (4), 689–705.

Bommer, W.H., Johnson, J.L., Rich, G.A., Podsakoff, P.M., & MacKenzie, S.B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology*, **48**, 587.

Bordin, E.S. (1994). Theory and research on the therapeutic working alliance. *Psychotherapy: Theory, Research and Practice*, **16**, 252–260.

Borman, W.C. (1997). 360° ratings: An analysis of assumptions and a research agenda for evaluating their validity. *Human Resource Management Review*, 7, 299, 315.

Borman, W.C., Buck, D.E., Hanson, M.A., Motowidlo, S.J., Stark, S., & Drasgow, F. (2001). An examination of the comparative reliability, validity and accuracy of

performance ratings made using computerized adaptive rating scales. *Journal of Applied Psychology*, **86**, 965–973.

Boswell, W.R. & Boudreau, J.W. (2002). Separating the developmental and evaluative performance appraisal uses. *Journal of Business and Psychology*, **16**, 391–412.

Bowen, C.C., Swim, J.K., & Jacobs, R.R. (2000). Evaluating gender biases on actual job performance of real people: A meta-analysis. *Journal of Applied Social Psychology*, **30**, 2194–2215.

Bowlby, J. (1979). *The making and breaking of affectional bonds*. London: Tavistock.

Bowman, J.S. (1999). Performance appraisal: Verisimilitude trumps veracity. *Public Personnel Management*, **28**, 557–594.

Bozeman, D.P. (1997). Interrater agreement in multi-source performance appraisal: A commentary. *Journal of Organizational Behavior*, **18**, 313–316.

Bracken, D.W., Timmreck, C.W., & Church, A.H. (Eds) (2001). *The handbook of multisource feedback*. San Francisco: Jossey-Bass.

Brett, J.F. & Atwater, L.E. (2001). 360° feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, **86**, 930–942.

Brown, T.C. (2003). The effect of verbal self-guidance training on collective efficacy and team performance. *Personnel Psychology*, **56**, 935–964.

Brown, T.C. & Latham, G.P. (2002). The effects of behavioural outcome goals, learning goals, and urging people to do their best on an individual's teamwork behaviour in a group problem solving task. *Canadian Journal of Behavioural Science*, **34**, 276–285.

Brown, T.C. & Latham, G.P. (in press). The effect of training in verbal self-guidance on performance effectiveness in a MBA program. *Canadian Journal of Behavioural Science*.

Brutus, S. & Derayeh, M. (2002). Multisource assessment programs in organizations: An insider's perspective. *Human Resource Development Quarterly*, **13**, 187–202.

Campbell, J.D., Garfinkel, R.B., & Moses, L. (1996). Strategies for success in measuring performance. *HR Magazine*, **41**, 98.

Cardy, R.L. & Keefe, T.J. (1994). Observational purpose and evaluative articulation in frame-of-reference training: The effects of alternative processing modes on rating accuracy. *Organizational Behavior and Human Decision Processes*, **57**, 338–357.

Cawley, B.D., Keeping, L.M., & Levy, P.E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, **83**, 615–631.

Cederblom, D. & Permerl, D.E. (2002). From performance appraisal to performance management. One agency's experience. *Public Personnel Management*, **31**, 131–140.

Church, A.H. (2000). Do higher performing managers actually receive better ratings? A validation of multirater assessment methodology. *Consulting Psychology Journal: Practice and Research*, **52**, 99–116.

Cialdini, R.B. (2001). The science of persuasion. *Scientific American*, **284**, 76–81.

Claus, L. (2003). Similarities and differences in human resource management in the European Union. *Thunderbird International Business Review*, **45**, 729–755.

Cohen, S.G. & Bailey, D.E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, **23**, 239–290.

Conway, J.M. (1996). Analysis and design of multitrait–multirater performance appraisal studies. *Journal of Management*, **22**, 139–162.

Conway, J.M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology*, **84**, 3–13.

Cronshaw, S.F. (1998). Job analysis: Changing nature of work. *Canadian Psychology*, **39**, 5–13.

DeNisi, A.S. & Kluger, A.N. (2000). Feedback effectiveness: Can 360-degree appraisals be improved? *Academy of Management Executive*, **14**, 129–139.

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**, 717–737.

DeVoe, S.E. & Iyengar, S.S. (2003). Managers' theories of subordinates: A cross-cultural examination of manager perceptions of motivation and appraisal of performance. *Organizational Behavior and Human Decision Processes*, **93**, 47–61.

Dewberry, C. (2001). Performance disparities between whites and ethnic minorities: Real differences of assessment bias? *Journal of Occupational and Organizational Psychology*, **74**, 656–673.

Dickerson, C.A., Thibodeau, R., Aronson, E., & Miller, D. (1992). Using cognitive dissonance to encourage water conservation. *Journal of Applied Social Psychology*, **22**, 841–854.

Dominick, P.G., Reilly, R.R., & McGourty, J.W. (1997). The effects of peer feedback on team member behaviour. *Group and Organization Management*, **22**, 508–520.

Dorner, D. (1991). The investigation of action regulation in uncertain and complex situations. In J. Rassmussen, G. Brehmer, & J. Leplat (Eds), *Distributed decision making: Cognitive models for cooperative work*. New York: John Wiley & Sons.

Druskat, V.U. & Wolff, S.B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology*, **84**, 58–74.

Dweck, C.S. (1999). *Self theories: Their role in motivation, personality, and development*. Philadelphia: Psychology Press.

Eagly, A.H., Karau, S.J., & Makhijani, M.G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, **117**, 125–145.

Eagly, A.H., Makhijani, M.G., & Klonsky, B.G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, **111**, 3–22.

Ellinger, A.D., Watkins, K.E., & Bostrom, R.P. (1999). Managers as facilitators of learning in learning organizations. *Human Resource Development Quarterly*, **10**, 105–125.

Ellis, A. (1999). Early theories and practices of rational emotive behaviour therapy and how they have been augmented and revised during the last three decades. *Journal of Rational-Emotive and Cognitive Behavior Therapy*, **17**, 69–93.

Emery, F.E. & Thorsrud, E. (1976). *Democracy at work*. Leiden, The Netherlands: Martinus Nijhoff.

Erdogan, B., Kraimer, M.L., & Liden, R.C. (2001). Procedural justice as a two-dimensional construct: An examination in the performance appraisal account. *Journal of Applied Behavioral Science*, **37** (2), 205–222.

Erez, M. (1994). Toward a model of cross-cultural industrial and organizational psychology. In H.C. Triandis & M.D. Dunnette (Eds), *Handbook of industrial and organizational psychology* (Vol. 4, pp. 559–607). Palo Alto, CA: Consulting Psychologists Press.

Erez, M.E., Kaplan, E., & Van-Dijk, D. (2004). *Goal setting and regulatory focus: An integrative approach*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Chicago.

Erikson, A. & Allen, T. (2003). *Linking 360° feedback to business outcome measures*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.

Facteau, J.D. & Craig, S.B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology*, **86**, 215–227.

Feild, H.S. & Holley, W.H. (1982). The relationship of performance appraisal system characteristics to verdicts in selected employment discrimination cases. *Academy of Management Journal*, **25**, 392–406.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, **7**, 117–140.

Fitzsimons, G. & Shah, J.Y. (2005). Instrumentality yours: How goal pursuits automatically affect social relations. Unpublished manuscript, Stanford University, Stanford, CA.

Fletcher, C. (1997). Self-awareness—A neglected attribute in selection and assessment? *International Journal of Selection and Assessment*, **5**, 183–187.

Fletcher, C. (2001). Performance appraisal and management: The developing research agenda. *Journal of Occupational and Organizational Psychology*, **74**, 473–487.

Fletcher, C. & Perry, E.L. (2001). Performance appraisal and feedback: A consideration of national culture and a review of contemporary research and future trends. In N.D. Anderson et al. (Eds), *Handbook of industrial, work and organizational psychology.* London: Sage.

Fox, S., Ben-Nahum, Z., & Yinnon, Y. (1989). Perceived similarity and accuracy of peer ratings. *Journal of Applied Psychology*, **74**, 781–786.

Fraley, R.C., Garner, J.P., & Shaver, P.R. (2000). Adult attachment and the defensive regulation of attention and memory: Examining the role of preemptive and postemptive defensive processes. *Journal of Personality and Social Psychology*, **79**, 816–826.

Freid, Y., Levi, A.S., Ben-David, H.A., Tiegs, R.B., & Avital, N. (2000). Rater positive and negative mood predispositions as predictors of performance ratings of ratees in simulated and real organizational settings. *Journal of Organizational and Occupational Psychology*, **73**, 373–378.

Frese, M. (2005). Grand theories and midrange theories. Cultural effects on theorizing and the attempt to understand active approaches to work. In K.G. Smith & M. Hitt (Eds), *The Oxford handbook of management theory: The process of theory development.* Oxford, UK: Oxford University Press.

Frese, M. & Zapf, D. (1994). Action as the core of work psychology: A German approach. In H.C. Triandis, M.D. Dunnette, & L.M. Hough (Eds), *Handbook of industrial and organizational psychology* (2nd edn). Palo Alto, CA: Consulting Psychologists Press.

Funderburg, S.A. & Levy, P.E. (1997). The influence of individual and contextual variables on 360-degree feedback system attitudes. *Group and Organization Management*, **22** (2), 210–235.

Gabris, G.T. & Ihrke, D.M. (2001). Does performance appraisal contribute to heightened levels of employee burnout? The results of one study. *Public Personnel Management*, **30**, 157–172.

General Electric Company (2000). *GE 2000 Annual Report*.

Ghorpade, J. (2000). Managing five paradoxes of 360-degree feedback. *Academy of Management Executive*, **14**, 140–151.

Gilbert, G.R., Hannan, E.L., & Lowe, K.B. (1998). Is smoking stigma clouding the objectivity of employee performance appraisal? *Public Personnel Management*, **27** (3), 285–300.

Goodman, J.S., Wood, R.E., & Hendrickx, M. (2004). Feedback specificity, exploration, and learning. *Journal of Applied Psychology*, **89**, 248–262.

Gordon, R.A. & Arvey, R.D. (2004). Age bias in laboratory and field settings: A meta-analytic investigation. *Journal of Applied Social Psychology*, **34**, 468–492.

Greenberg, J. (2000). Promote procedural justice to enhance acceptance of work outcomes. In E. Locke (Ed.), *Handbook of principles of organizational behavior.* Oxford, UK: Blackwell.

Greguras, G.J. & Robie, C. (1998). A new look at within-source interrater reliability of 360-degree feedback ratings. *Journal of Applied Psychology*, **83**, 960–968.

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

Hall, D.T., Otazo, K.L., & Hollenbeck, G.P. (1999). Behind closed doors: What really happens in executive coaching. *Organizational Dynamics*, **27**, 39–53.

Halperin, K., Snyder, C.R., Shenkel, R.J., & Houston, B.K. (1976). Effects of source status and message favorability on acceptance of personality feedback. *Journal of Applied Psychology*, **61**, 85–88.

Head, T., Haug, R., Krabbenhoft, A., & Ma, C. (2000). Issues and trends in international business law: Implications for OD consultants. *Organization Development Journal*, **18**, 62–73.

Hedge, J.W. & Teachout, M.S. (2000). Exploring the concept of acceptability as a criterion for evaluating performance measures. *Group and Organization Management*, **25**, 22–44.

Heslin, P. & Latham, G.P. (2004). The effect of upward feedback on managerial behavior. *Applied Psychology: An International Review*, **53**, 23–38.

Heslin, P.A., Latham, G.P., & VandeWalle, D. (2005). Implicit person theory and performance appraisal. *Journal of Applied Psychology*, **90**, 842–856.

Heslin, P.A., VandeWalle, D., & Latham, G.P. (2005). *Keen to help: Managers implicit person theory and their subsequent employee coaching*. Paper presented at the Annual Meeting of the Academy of Management, Honolulu.

Higgins, E.T. (1999). Persons or situations: Unique explanatory principles or variability in general principles? In D. Cervone & Y. Shoda (Eds), *The coherence of personality: Social cognitive biases of consistency, variability and organization* (pp. 61–93). New York: Guilford Press.

Hovland, C. & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, **15**, 635–650.

Jain, H.C., Sloan, P.J., & Horowitz, F.M. (2003). *Employment equity and affirmative action: An international comparison.* Armonk, NY: Sharpe.

Jawahar, I.M. (2001). Attitudes, self-monitoring and appraisal behaviors. *Journal of Applied Psychology*, **86**, 875–883.

Jelley, R.B. & Goffin, R.D. (2001). Can performance-feedback accuracy be improved? Effects of rater priming and rating-scale format on rating accuracy. *Journal of Applied Psychology*, **86**, 134–144.

Johnson, D.E., Erez, A., Kiker, D.S., & Motowidlo, S.J. (2002). Liking and attributions of motives as mediators of the relationships between individuals' reputations, helpful behaviors and raters' reward decisions. *Journal of Applied Psychology*, **87**, 808–815.

Kalinauckas, P. & King, H. (1995). Coaching gets to parts that other methods cannot reach. *People Management*, **1**, 51.

Kane, J.S., Bernardin, H.J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: Three studies. *Academy of Management Journal*, **38**, 1036–1051.

Kaplan, R.S. & Norton, D.P. (1993). Putting the balanced scorecard to work. *Harvard Business Review*, Sept–Oct, 134–147.

Keown-Gerrard, J.A. & Sulsky, L.M. (2001). The effects of task information training and frame-of-reference training with situational constraints on rating accuracy. *Human Performance*, **14**, 305–320.

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**, 885–896.

Kleingeld, A., Van Tuijl, H., & Algera, J.A. (2004). Participation in the design of performance management systems: A quasi-experimental field study. *Journal of Organizational Behavior*, **25**, 831–851.

Kluger, A.N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, **119**, 254–284.

Komaki, J.L. & Desselles, M.L. (1994). *Supervision reexamined: The role of monitors and consequences*. Boston: Allen & Unwin.

Korsgaard, M.A., Meglino, B.M., & Lester, S.W. (2004). The effect of other orientation on self-supervisor rating agreement. *Journal of Organizational Behavior*, **25**, 873–891.

Korsgaard, M.A., Roberson, L., & Rymph, R.D. (1998). What motivates fairness? The role of subordinate assertive behavior on manager's interactional fairness. *Journal of Applied Psychology*, **83** (5), 731–744.

Kraiger, K. & Ford, J.K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology*, **70**, 56–65.

Lam, S.S.K. & Schaubroeck, J. (1999). Total quality management and performance appraisal: An experimental study of process versus results and group versus individual approaches. *Journal of Organizational Behavior*, **20**, 445–457.

Lam, S.S.K., Yik, M.S.M., & Schaubroeck, J. (2002). Responses to formal performance appraisal feedback: The role of negative affectivity. *Journal of Applied Psychology*, **87**, 192–201.

Lance, C.E. (1994). Test of a latent structure of performance ratings derived from Wherry's (1952) theory of ratings. *Journal of Management*, **20**, 757–771.

Landau, J. (1995). The relationship of race and gender to managers' ratings of promotion potential. *Journal of Organizational Behavior*, **16**, 391–400.

Langan-Fox, J., Waycott, J., Morizzi, M., & McDonald, L. (1998). Predictors of participation in performance appraisal: A voluntary system in a blue-collar work environment. *International Journal of Selection and Assessment*, **6** (4), 249–260.

Latham, G.P. (1986). Job performance and appraisal. In C. Cooper & I. Robertson (Eds), *Review of industrial and organizational psychology*. Chichester, UK: John Wiley & Sons.

Latham, G.P. (2001). The importance of understanding and changing employee outcome expectancies for gaining commitment to an organizational goal. *Personnel Psychology*, **54**, 707–716.

Latham, G.P. & Budworth, M. (2005, April). *Anchoring the subsequent appraisal of others on one's own performance appraisal*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Los Angeles, California.

Latham, G.P. & Heslin. P. (2003). Training the trainee as well as the trainer: Lessons to be learned from clinical psychology. *Canadian Psychology*, **44**, 218–231.

Latham, G.P. & Latham, S.D. (2000). Overlooking theory and research in performance appraisal at one's peril: Much done, more to do. In C. Cooper & E.A. Locke (Eds), *International review of industrial and organizational psychology*. Chichester, UK: John Wiley & Sons.

Latham, G.P., Locke, E.A., & Fassina, N.E. (2002). The high performance cycle: Standing the test of time. In S. Sonnentag (Ed.), *The psychological management of individual performance: A handbook in the psychology of management in organizations* (pp. 201–228). Chichester, UK: John Wiley & Sons.

Latham, G.P. & McCauley, C.D. (2005). Leadership in the private sector: Yesterday versus tomorrow. In C. Cooper (Ed.), *The twenty-first century manager*. Oxford, UK: Oxford University Press.

Latham, G.P. & Pinder, C.C. (2005). Work motivation theory and research at the dawn of the twenty-first century. *Annual Review of Psychology*, **56**, 485–516.

Latham, G.P. & Seijts, G.H. (1997). The effect of appraisal instrument on managerial perceptions of fairness and satisfaction with appraisals from their peers. *Canadian Journal of Behavioural Science*, **29**, 275–282.

Latham, G.P. & Seijts, G.H. (1999). The effects of proximal and distal goals on performance on a moderately complex task. *Journal of Organizational Behavior*, **20**, 421–429.

Latham, G.P., Skarlicki, D., Irvine, D., & Siegel, J. (1993). The increasing importance of performance appraisals to employee effectiveness in organizational settings in North America. In C. Cooper & I. Robertson (Eds), *International review of industrial and organizational psychology* (pp. 87–132). Chichester, UK: John Wiley & Sons.

Latham, G.P. & Wexley, K.N. (1994). *Increasing productivity through performance appraisal* (2nd edn). Reading, MA: Wesley.

Latham, G.P., Wexley, K.N., & Pursell, E.D. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology*, **60**, 550–555.

Lefkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational and Organizational Psychology*, **73**, 67–85.

Lefkowitz, J. & Battista, M. (1995). Potential sources of criterion bias in supervisor ratings used for test validation. *Journal of Business and Psychology*, **9**, 389–414.

LePine, J.A., Erez, A., & Johnson, D.E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology*, **87**, 52–65.

Levy, P.E. & Williams, J.R. (1998). The role of perceived system knowledge in predicting appraisal reactions, job satisfaction and organizational commitment. *Journal of Organizational Behavior*, **19**, 53–65.

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: A 35-year odyssey. *American Psychologist*, **57**, 705–717.

London, M. & Smither, J.W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. *Personnel Psychology*, **48**, 803–839.

London, M., Smither, J.W., & Adsit, D.J. (1997). Accountability: The Achilles' heel of multisource feedback. *Group and Organization Management*, **22**, 162–184.

Luthans, F. & Peterson, S.J. (2003). 360-degree feedback with systematic coaching: empirical analysis suggests a winning combination. *Human Resource Management*, **42**, 243–256.

Mann, S.L. & Budworth, M.H. (2005). *Self and peer rater agreement of counterproductive performance*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Los Angeles.

Markus, H.R. & Kitayama, S. (1991). Culture and the self: implications for cognition, emotion and motivation. *Psychological Review*, **98**, 224–253.

Martin, D., Bartol, K., & Kehoe, P. (2000). The legal ramifications of performance appraisal: The growing significance. *Public Personnel Management*, **29**, 379–406.

Maurer, T.J., Mitchell, D.R.D., & Barbeite, F.G. (2002). Predictors of attitudes toward a 360° feedback system and involvement in post-feedback management development activity. *Journal of Occupational and Organizational Psychology*, **75**, 87–107.

Mayer, R.C. & Davis, J.H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, **84**, 123–136.

Meichenbaum, D. (1977). *Cognitive behavior modification: An integrative approach.* New York: Plenum Press.

Miller, J.S. & Cardy, R.L. (2000). Self-monitoring and performance appraisal: Rating outcomes in project teams. *Journal of Organizational Behavior*, **21**, 609–626.

Millman, Z. & Latham, G.P. (2001). Increasing reemployment through training in verbal self-guidance. In M. Erez, U. Kleinbeck, & H. Thierry (Eds), *Work motivation in the context of a globalizing economy*. Mahwah, NJ: Lawrence Erlbaum.

Murphy, K.R. & Cleveland, J.N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.

Nease, A.A., Mudgett, B.O., & Quinones, M.A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of Applied Psychology*, **84**, 806–814.

Noonan, L.E. & Sulsky, L.M. (2001). Impact of frame-of-reference training on alternative training effectiveness criteria in a Canadian military sample. *Human Performance*, **14**, 3–26.

Northcraft, G.B. & Earley, P.C. (1989). Technology, credibility, and feedback use. *Organizational Behavior and Human Decision Processes*, **44**, 83–96.

O'Leary-Kelly, A., Martocchio, J., & Frink, D. (1994). A review of the influence of group goals on group performance. *Academy of Management Journal*, **37**, 1285–1301.

Olivero, G., Bane, K.D., & Kopelman, R.E. (1997). Executive coaching as a transfer of training tool: Effects on productivity in a public agency. *Public Personnel Management*, **26**, 461–469.

Organ, D. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, **10**, 85–98.

Ostroff, C., Atwater, L.E., & Feinberg, B.J. (2004). Understanding self–other agreement: A look at rater and ratee characteristics, context and outcomes. *Personnel Psychology*, **57**, 333–375.

Pazy, A. & Oron, I. (2001). Sex proportion and performance evaluation among high-ranking military officers. *Journal of Organizational Behavior*, **22**, 689–702.

Pelletier, L.G. & Vallerand, R.J. (1996). Supervisors' beliefs and subordinates' intrinsic motivation: A behavioral confirmation analysis. *Journal of Personality and Social Psychology*, **71**, 331–340.

Pettijohn, C.E., Pettijohn, L.S., & d'Amico, M. (2001). Characteristics of performance appraisals and their impact on sales force satisfaction. *Human Resource Development Quarterly*, **12**, 127–146.

Pratkanis, A. & Aronson, E. (2001). *Age of propaganda: The everyday use and abuse of persuasion*. New York: Freeman.

Reilly, R.R., Smither, J.W., & Vasilopoulos, N.L. (1996). A longitudinal study of upward feedback. *Personnel Psychology*, **49**, 599–612.

Renn, R.W. (2003). Moderation by goal commitment of the feedback–performance relationship: Theoretical explanation and preliminary study. *Human Resource Management Review*, **13**, 561–580.

Richardson, L. (1998). Five-minute sales coaching. *Training and Development*, **52**, 53–57.

Rotundo, M. & Sackett, P.R. (2002). The relative importance of task, citizenship and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, **87**, 66–80.

Saks, A.M. & Waldman, D.A. (1998). The relationship between age and job performance evaluations for entry-level professionals. *Journal of Organizational Behavior*, **19**, 409–419.

Salter, J.S. (1996). Officer attitude and opinion survey: Career management and job performance issues. *Australian Army First Psychological Research Unit Report*, May, 20.

Schleicher, D.J. & Day, D.V. (1998). A cognitive evaluation of framework-of-reference rater training: Content and process issues. *Organizational Behavior and Human Decision Processes*, **73**, 76–101.

Scholz, U., Dona, B.G., Sud, S., & Schwarzer R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment*, **18**, 242–251.

Schrader, B.W. & Steiner, D.D. (1996). Common comparison standards: An approach to improving agreement between self and supervisory performance. *Journal of Applied Psychology*, **81**, 813–820.

Scullen, S.E., Bergey, P.K., & Aiman-Smith, L. (2005). Forced distribution rating systems and their improvement of workforce potential: A baseline simulation. *Personnel Psychology*, **58**, 1–32.

Scullen, S.E., Mount, M.K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, **85** (6), 956–970.

Seifert, C.F., Yukl, G., & McDonald, R.A. (2003). Effects of multi-source feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *Journal of Applied Psychology*, **88**, 561–569.

Seijts, G.H. & Latham, G.P. (2001). The effect of learning, outcome, and proximal goals on a moderately complex task. *Journal of Organizational Behavior*, **22**, 291–307.

Seijts, G.H. & Latham, G.P. (2005) Learning versus performance goals: When should each be used? *Academy of Management Executive*, **19**, 124–131.

Seijts, G.H., Latham, G.P., Tasa, K., & Latham, B.W. (2004). Goal setting and goal orientation: An integration of two different yet related literatures. *Academy of Management Journal*, **47**, 227–239.

Seligman, M.E.P. (1998). *Learned optimism.* New York: Pocket Books.

Shah, J. (2003a). Automatic for the people: How representations of significant others implicitly affect goal pursuit. *Journal of Personality and Social Psychology*, **84**, 661–681.

Shah, J. (2003b). The motivational looking glass: How significant others implicitly affect goal appraisals. *Journal of Personality and Social Psychology*, **85**, 424–439.

Shah, J.Y. (2005). The automatic pursuit and management of goals. *Current Directions in Psychological Science*, **14**, 10–13.

Smither, J.W., London, M., Flautt, R., Vargas, Y., & Kucine, I. (2003). Can working with an executive coach improve multisource feedback ratings over time? A quasi-experimental field study. *Personnel Psychology*, **56**, 23–44.

Smither, J.W., London, M., & Reilly, R.R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology*, **58**, 33–66.

Smither, J.W., London, M., Vasilopoulos, N.L., Reilly, R.R., Millsap, R.E., & Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology*, **48**, 1–34.

Smither, J.W. & Walker, A.G. (2004). Are the characteristics of narrative comments related to improvement in multirater feedback ratings over time? *Journal of Applied Psychology*, **89**, 575–581.

Steers, R.M. & Sanchez-Runde, C.J. (2002). Culture, motivation, and work behavior. In M.J. Gannon and K.L. Newman (Eds), *The Blackwell handbook of principles of cross-cultural management*. Bodmin, UK: MPG Books.

Stone, J., Aronson, E., Crain, A.L., Winslow, M.P., & Fried, C.B. (1994). Inducing hypocrisy as a means of encouraging young adults to use condoms. *Personality and Social Psychology Bulletin*, **20**, 116–128.

Strauss, J.P., Barrick, M., & Connerley, M. (2001). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and liking. *Journal of Occupational and Organizational Psychology*, **74** (5), 637–657.

Sue-Chan, C. & Latham, G.P. (2004). The relative effectiveness of external, peer and self-coaches. *Applied Psychology: An International Review*, **53**, 260–278.

Sulsky, L.M. & Keown, J.L. (1998). Performance appraisal in the changing world of work: Implications for the meaning and measurement of work performance. *Canadian Psychology*, **39** (1–2), 52–59.

Taylor, M.S., Tracy, K.B., Renard, M.K., Harrison, J.K., & Carroll, S.J. (1995). Due process in performance appraisal: A quasi-experiment in procedural justice. *Administrative Science Quarterly*, **40**, 495–523.

Trist, E.L. (1977). Collaboration in work settings. *Journal of Applied Behavioral Science*, **13**, 268–278.

Tyler, K. (1997). Prepare managers to become career coaches. *HRMagazine*, **42**, 98–101.

Tziner, A. (1999). The relationship between distal and proximal factors and the use of political considerations in performance appraisal. *Journal of Business and Psychology*, **14**, 217–231.

Tziner, A., Joannis, C., & Murphy, K.R. (2000). A comparison of three models of performance appraisal with regard to goal properties, goal perception and ratee satisfaction. *Group and Organization Management*, **25**, 175–190.

Tziner, A. & Kopelman, R.E. (2002). Is there a preferred performance rating format? A non-psychometric perspective. *Applied Psychology: An International Review*, **51**, 479–503.

Tziner, A., Latham, G.P., Price, B.S., & Haccoun, R. (1996). Development and validation of a questionnaire for measuring perceived political considerations in performance appraisal. *Journal of Organizational Behavior*, **17**, 179–190.

Tziner, A., Murphy, K.R., & Cleveland, J.N. (2001). Relationships between attitudes towards organizations and performance appraisal systems and rating behaviors. *International Journal of Selection and Assessment*, **9**, 226–239.

Uggerslev, K.L. & Sulsky, L.M. (2002). Presentation modality and indirect performance information: Effects on ratings, reaction and memory. *Journal of Applied Psychology*, **87**, 940–950.

Van Scotter, J.R. & Motowidlo, S.J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, **81**, 525–531.

Varma, A., DeNisi, A.S., & Peters, L.H. (1996). Interpersonal affect and performance appraisal: A field study. *Personnel Psychology*, **49**, 341–361.

Walker, A.G. & Smither, J.W. (1999). A five-year study of upward feedback: What managers do with their results matters. *Personnel Psychology*, **52**, 393–423.

Watson Wyatt (2002). *Human Capital Index: Human capital as a lead indicator of shareholder value*. Washington, DC: Author.

Wayne, S.J. & Liden, R.C. (1995). Effects of impression management on performance ratings: A longitudinal study. *Academy of Management Journal*, **38**, 232–260.

Weisband, S. & Atwater, L. (1999). Evaluating self and others in electronic and face-to-face groups. *Journal of Applied Psychology*, **84** (4), 632–639.

Welch, J. & Byrne, J.A. (2001). *Jack: Straight from the gut*. New York: Watner Books.

Werner, J.M. & Bolino, M.C. (1997). Explaining US courts of appeals decisions involving performance appraisal: Accuracy, fairness, and validation. *Personnel Psychology*, **50**, 1–24.

Winters, D. & Latham, G.P. (1996). The effect of learning versus outcome goals on a simple versus a complex task. *Group and Organization Management*, **21**, 236–250.

Woehr, D.J. & Huffcutt, A.I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, **67**, 189–205.

Woodford, K. & Maes, J.D. (2002). Employee performance evaluations: Administering and writing them correctly in the multi-national setting. *Equal Opportunities International*, **21**, 1–8.

Zagona, S.V. & Harter, R. (1966). Credibility of source and recipient's attitude: Factors in the perception and retention of information on smoking behavior. *Perceptual and Motor Skills*, **23**, 155–168.

Zazanis, M.M., Zaccaco, S.J., & Kilcullen, R.N. (2001). Identifying motivation and interpersonal performance using peer evaluations. *Military Psychology*, **13**, 73–88.

Zhou, J. (1998). Feedback valence, feedback style, task autonomy, and achievement orientation: Interactive effects on creative performance. *Journal of Applied Psychology*, **83**, 261–276.