



Pergamon

Journal of Management 2004 30(6) 881–905

**JOURNAL OF
MANAGEMENT**

The Social Context of Performance Appraisal: A Review and Framework for the Future

Paul E. Levy*

*Department of Psychology, The University of Akron,
290 East Buchtel Avenue, Akron, OH 44325-4301, USA*

Jane R. Williams

*Psychology Department, Indiana University-Purdue University Indianapolis,
402 N. Blackford St., Indianapolis, IN 46142, USA*

Received 3 January 2004; received in revised form 24 May 2004; accepted 1 June 2004

Available online 14 July 2004

Performance appraisal research over the last 10 years has begun to examine the effects of the social context on the appraisal process. Drawing from previous theoretical work, we developed a model of this process and conducted a systematic review of the relevant research. This review of over 300 articles suggests that as a field we have become much more cognizant of the importance of the social context within which the performance appraisal process operates. First, research has broadened the traditional conceptualization of performance appraisal effectiveness to include and emphasize ratee reactions. Second, the influence that the feedback environment or feedback culture has on performance appraisal outcomes is an especially recent focus that seems to have both theoretical and applied implications. Finally, there appears to be a reasonably large set of distal variables such as technology, HR strategies, and economic conditions that are potentially important for understanding the appraisal process, but which have received very little research attention. We believe that the focus of recent performance appraisal research has widespread implications ranging from theory development and enhancement to practical application.

© 2004 Elsevier Inc. All rights reserved.

Organizations have been doing performance appraisal for many years and researchers have been investigating, studying, and trying to improve performance appraisal for almost as long (Farr & Levy, *in press*). We have learned a great deal about the performance appraisal process

* Corresponding author. Tel.: +1 330 972 8369; fax: +1 330 972 5174.

E-mail addresses: pelevy@uakron.edu (P.E. Levy), jrwillim@iupui.edu (J.R. Williams).

as a result of the research that has been conducted. Our purpose is to build a framework through which we can review the most current work in performance appraisal and make suggestions for future research.

The 1990s and the Beginnings of the Social Context

The landscape of performance appraisal changed dramatically from a purely measurement focus to one in which the cognitive processes of appraisal became paramount as a result of Landy and Farr's (1980) classic paper (see also Feldman, 1981). A second major change to this research area slowly evolved during the 1990s and this is the change that is the focus of our paper. In this section, we'd like to convey that new terrain by discussing a few key papers that either helped change or helped to articulate how that landscape was beginning to change.

Murphy and Cleveland published their first book on performance appraisal in 1991. In it they made a very strong case for needing a new model or approach to studying performance appraisal. They argued that previous models (e.g., Landy & Farr's, 1980) were very useful, but (1) paid inadequate attention to the organizational context in which appraisals occur, and (2) failed to demonstrate links between the appraisal research and appraisal practice (Murphy & Cleveland, 1991).

An outstanding paper by Ilgen, Barnes-Farrell and McKellin (1993) expanded on both of the points made by Murphy and Cleveland. First, they reviewed the cognitive process research of the 1980s and concluded that while it certainly improved our understanding of the appraisal process it had not provided enough information to practitioners regarding how better to do performance appraisal. Recall that Landy and Farr (1980) basically argued that we had gone far enough with the format research and needed to move in a different direction. Ilgen et al. go even one step further by stating the following: "We feel that it is far more likely that the major problems facing performance appraisals at this time lie *neither* (emphasis added) in the cognitive process domain nor in that of rating scale construction" (p. 361). They go on to argue convincingly that a potentially valuable focus for performance appraisal research to take is one that emphasizes the rating environment or the "social milieu" in which the participants in the appraisal process find themselves.

Finally, one important review article helped to highlight the progress of this change. Bretz, Milkovich and Read (1992) reviewed the published literature between 1985 and 1990. They noted that cognitive processing clearly dominated this period of performance appraisal research with an emphasis also on psychometrics as they related to halo error and accuracy. They highlighted some of the same issues that were targeted by Ilgen et al. (1993) regarding the gap between the cognitive and psychometric research of this period and the more practice-oriented issues involved in performance appraisal. Bretz et al. (1992) reported that:

The predominance of studies examined information processing and psychometric issues, yet virtually no systematic research exists on how the organizational context affects the rater/ratee relationship. (p. 330)

Bretz et al. concluded that in order to gain a better understanding of the appraisal process and if research is to better inform appraisal practice, a great deal more research attention must be paid to the effects of the performance appraisal context. Ilgen et al. (1993) proposed that research resources be redirected away from cognitive processing, which only accounts for a limited amount of variance in the appraisal process, and focused instead on other influences such as the social milieu or rating environment. Thus, our purpose here is to review the current literature to evaluate how well the recommendations of these scholars have been heeded.

The Social Context of Performance Appraisal

We argue that identifying, measuring, and defining the organizational context in which appraisal takes place is integral to truly understanding and developing effective performance appraisals. Further, we believe that this has been the framework driving the performance appraisal research since about 1990 and into the beginning of the 21st century. Whether it's discussed as the social-psychological process of performance appraisal (Murphy & Cleveland, 1991), the social context of performance appraisal (Ferris, Judge, Rowland & Fitzgibbons, 1994), the social milieu of performance appraisal (Ilgen et al., 1993), performance appraisal from the organizational side (Levy & Steelman, 1997), the games that raters and ratees play (Kozlowski, Chao & Morrison, 1998), or the due process approach to performance appraisal (Folger, Konovsky & Cropanzano, 1992), we argue along with these other scholars that performance appraisal takes place in a social context and that context plays a major role in the effectiveness of the appraisal process and how participants react to that process (Farr & Levy, *in press*).

It has been suggested elsewhere that research over the last 10 years has moved noticeably away from a limited psychometric scope and toward an emphasis on variables that compose the social context (Fletcher, 2001). Although we agree with this conclusion and think that a new backdrop has emerged, no specific analysis, review, or synthesis of the literature has been attempted to validate this conclusion. In attempting to bring this kind of analysis to the literature we began by developing a streamlined model denoting the potential role played by the social context in the appraisal process. This model, adapted and expanded from Murphy and Cleveland (1991, 1995), is depicted in Figure 1 and its purpose is to serve as a heuristic framework that guides the remainder of our paper and, potentially, future research endeavors (Figure 1).

Our next step was to embark on a thorough review of the performance appraisal literature with a very clear emphasis on the social context as we have broadly defined it in Figure 1. To do this, we conducted a series of computerized searches of the published literature in performance appraisal (PsycINFO and Business Source Premier) from 1995 to 2003. Although 1995–2003 is our focus to keep the database somewhat manageable, we actually conducted searches back to 1990 and found about 600 published articles. We draw on some of the pre-1995 work in our discussions, but will limit ourselves mostly to the period of 1995–2003 for which we uncovered about 360 published articles. Although this survey of the literature is not exhaustive because it focused chiefly on research published in journals and does not include unpublished papers (e.g., dissertations and conference papers) nor

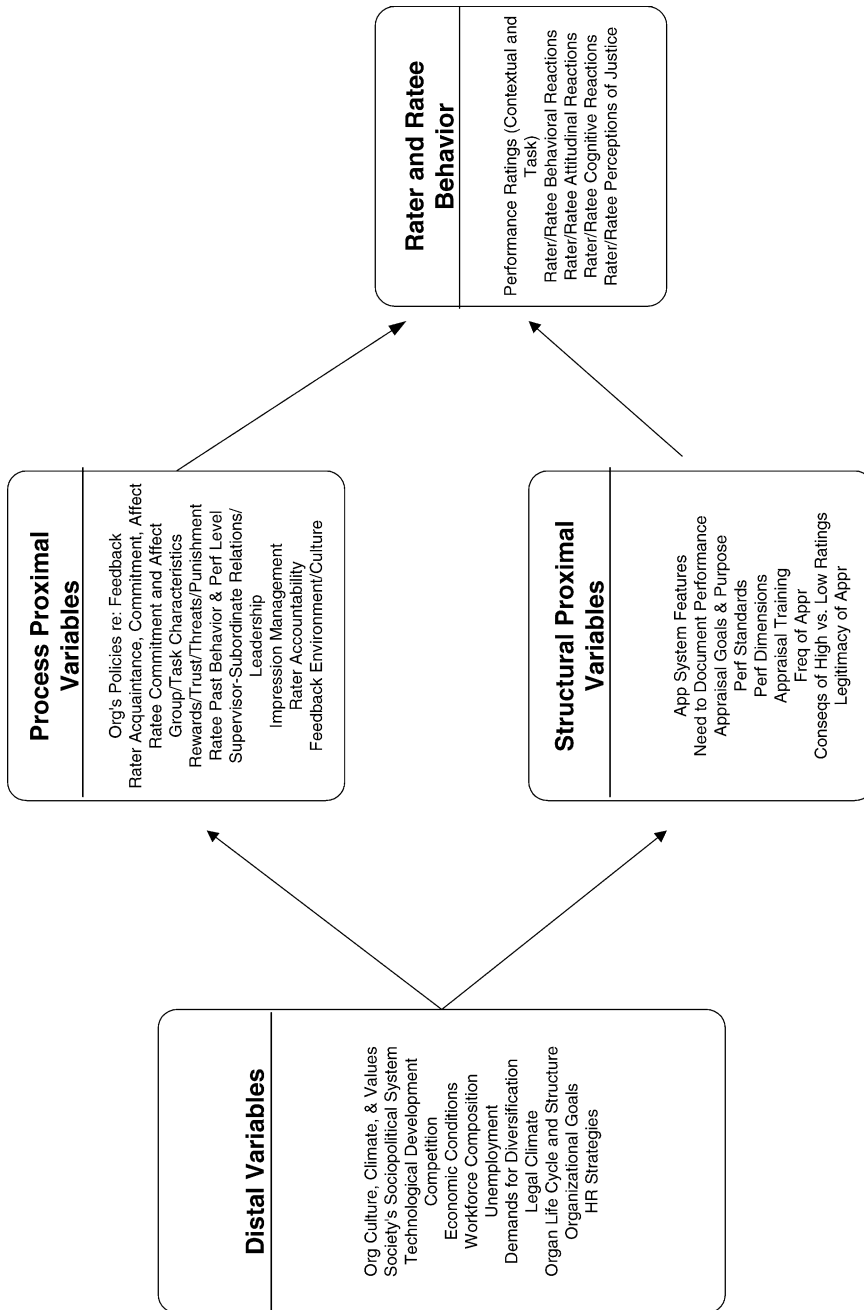


Figure 1. The social context of performance appraisal.

does it sample fully from the many book chapters that were published during this time, we do feel that it provides an accurate and representative snapshot of the areas on which scholars have chosen to focus their research attention and resources.

Performance Appraisal Literature Review, 1995–2003

Distal Factors

Our definition of distal variables is generally consistent with [Murphy and Cleveland \(1995\)](#) and others ([Bretz et al., 1992](#); [Harris, 1994](#)). Specifically, distal variables are broadly construed as contextual factors that affect many human resource systems, including performance appraisal. In other words, distal variables are not necessarily related only to performance appraisal, but they may have unique effects on the performance appraisal process that are useful to understand and consider.

Distal factors include but are not limited to organizational climate and culture, organizational goals, human resource strategies, external economic factors, technological advances, and workforce composition. We believe these factors have an effect on rater and ratee behavior, although not directly. For instance, an organization that espouses a continuous learning culture may structure and implement a very different type of performance appraisal system than an organization without such a culture. Also, the HR strategies adopted by an organization ([Harris, 1994](#)) may have an effect on the type of appraisal system adopted by an organization (e.g., developmental vs. administrative).

A review of the performance appraisal literature over the last 7–10 years reveals little systematic empirical work on the distal variables listed in [Figure 1](#) other than a bit on culture, climate, and technology issues (see, e.g., [Hebert & Vorauer, 2003](#); [Miller, 2003](#)). While this is at some levels disappointing, it is rather understandable. First, there is little theory specific to performance appraisal to methodically guide this level of research. Second, the breadth of the constructs we construe as distal make it difficult to measure and implement within a research setting. Third, given the distal nature of these factors, their direct effects on performance appraisal behavior may be small. Perhaps closer examination of the relationships between distal and proximal relationships would prove more fruitful. Even with the difficulties regarding this type of research, however, we believe it will be important to continue examining these factors to fully understand the social context in which performance appraisal operates. Given the relative lack of research on distal compared to proximal factors, however, the current review will focus on research that has considered proximal variables.

Process Proximal Variables

The next two sections of the paper will underscore those proximal variables (both process and structural) receiving attention in the recent appraisal literature. In organizing our literature review and trying to make sense out of the hundreds of studies published since 1995, we chose to categorize the proximal variables as either *process* (i.e., having direct impact on how the appraisal process is conducted including things such as accountability

or supervisor–subordinate relationships), or *structural* (i.e., dealing with the configuration or makeup of the appraisal itself and including things like the appraisal dimensions or frequency of appraisal).

Rater issues. Since 1995, researchers have shown considerable interest in variables related to the individual doing the appraisal. Our literature review identified *rater affect* as one of the most studied rater variables. Although the literature has not been consistent regarding a formal definition of affect in performance appraisal (Lefkowitz, 2000), a good general definition linked to most of this research involves liking or positive regard for one's subordinate. The Affect Infusion Model (Forgas & George, 2001) suggests that affective states impact on judgments and behaviors and, in particular, affect or mood plays a large role when tasks require a degree of constructive processing. For instance, in performance appraisal, raters in good moods tend to recall more positive information from memory and appraise performance positively (Sinclair, 1988). Consistent with the Affect Infusion Model, a few recent studies have examined the role of mood or affect in performance appraisal. Lefkowitz (2000) conducted a thorough and important review of affective regard and performance appraisal, finding over 20 studies in all and about 13 between 1990 and 1998. In summarizing these findings he reported that affective regard is related frequently to higher appraisal ratings, less inclination to punish subordinates, better supervisor–subordinate relationships, greater halo, and less accuracy. Lefkowitz also developed a model of the interpersonal determinants of performance appraisal that seems to hold great value for providing a framework to guide future research in this area.

A couple of recent studies have looked at the role of similarity in personality (Bates, 2002; Strauss, Barrick & Connerley, 2001) and similarity in affect levels between raters and ratees, finding that similarity is related to appraisal ratings. Antonioni and Park (2001) found that affect was more strongly related to rating leniency in upward and peer ratings than it was in traditional top-down ratings and that this effect was stronger when raters had observational time with their subordinates. They concluded from this that raters pay so much attention to their positive regard for subordinates that increased observations result in noticing (or constructing) more specific behaviors that fit their affect-driven schema.

In some of their work, DeNisi and his colleagues (Robbins & DeNisi, 1998; Varma, Denisi & Peters, 1996) have found that although affect is positively related to appraisal ratings, it is more strongly related to more subjective trait-like ratings, than to ostensibly more objective task-based ratings. Further, keeping performance diaries (DeNisi & Peters, 1996; Varma & Stroh, 2001) tended to increase the strength of that relationship between affect and performance ratings leading the authors to conclude that perhaps affect follows from subordinate performance level rather than the other way around. Allen and Rush (1998), in their study looking at the relationship between OCBs and task performance, uncovered a similar finding in that affect served as a mediator of the OCB—task performance relationship. This too suggests that affect may not always be a bias in the ratings process as was traditionally believed, but may simply result from performance levels or, in this case, from contextual performance. That is, perhaps high performing subordinates are liked more because they are high performers.

A second broad area related to raters that has received considerable research attention has to do with the *motivation* of the raters. Traditionally, research seemed to assume that

raters were motivated to rate accurately. In particular, the cognitive research of the 1980s stemming from Landy and Farr's (1980) call for process research approached the problem in this way with the assumption that (1) raters were motivated to rate accurately and (2) the problems in the appraisal process involved cognitive processing errors and complexities. More recently, however, researchers have begun to question whether all or even most raters are truly motivated to rate accurately. Some of this impetus seems to be in response to practitioner demands for research that informs their practice (for a review, see Ilgen et al., 1993). This has led to research which has attempted to identify and understand other elements of rater's motivation and how that motivation affects the appraisal process.

One line of research related to rater's motivation has focused on the role of *individual differences* and *rating purpose* on rating leniency. Most practitioners report overwhelming leniency on the part of their raters and this rating elevation has been found in empirical papers as well as surveys of organizations (Murphy & Cleveland, 1995). Villanova, Bernardin, Dahmus and Sims (1993) developed a measure to tap the extent to which raters were uncomfortable doing performance appraisal. They reported that individuals who were higher on this scale (Performance Appraisal Discomfort Scale (PADS)) were also more likely to give elevated ratings because they didn't want to deal with the discomfort and conflict that often comes with delivering negative feedback. In a second study, Bernardin and his colleagues (Bernardin, Cooke & Villanova, 2000) demonstrated that raters who were high on the Big Five factor of Agreeableness and low on Conscientiousness were those most likely to provide elevated ratings. These findings corroborated the notions that high Agreeableness individuals are cooperative, trustful, and sympathetic in nature while high Conscientiousness individuals are focused on excellence, very careful, and quite thorough.

The role of *attributions* in the performance appraisal process has also attracted some recent research attention. In some of these studies investigators have examined how the attributions that raters make for ratees' behaviors affect their motivation to rate or their actual rating. For instance, using a traditional social psychological framework, Struthers, Weiner and Allred (1998) found that whether individuals opted for consoling, reprimanding, transferring, demoting, or firing a hypothetical employee depended in large part on the extent to which the rater believed that the exhibited behavior was due to ability or effort. In a related vein, Johnson, Erez, Kiker and Motowidlo (2002) found that both liking and attributions mediated the relationships between reputation and reward decisions. More specifically, raters consider ratees' behaviors and their reputations when drawing attributional inferences and deciding on appropriate rewards. The implications of this line of research are clear: attributional processing is an important element of the rating process and these attributions, in part, determine raters' reactions and ratings.

A second line of research related to rater motivation has to do with *rater accountability*, which is the perceived potential to be evaluated by someone and being held responsible for one's decisions or behaviors (Fried, Tieg & Bellamy, 1992; Frink & Ferris, 1998). With respect to performance appraisal, accountability is typically thought of as the extent to which a rater is held answerable to someone else for his or her ratings of another employee. In one of the earlier studies on accountability, Klimoski and Inks (1990) reported that raters distorted appraisal ratings more when they were to be held accountable to the ratee for those ratings. For instance, when participants anticipated a face-to-face feedback meeting with a poor performing ratee, they rated the ratee more favorably than did other participants

who did not expect a face-to-face meeting. They concluded that accountability can result in distortions of performance ratings. Mero and Motowidlo (1995) demonstrated that raters told that ratees had been rated too low in the past responded by inflating ratings while others told that they would have to defend their ratings in writing provided more accurate ratings. In a recent follow up to this study (Mero, Motowidlo & Anna, 2003), it was hypothesized that the accountability pressures on raters to justify ratings may operate through an increased motivation to better prepare themselves for their rating task. This was manifested in raters paying more attention to performance and recording better performance-related notes. A related study looking at accountability forces in performance appraisal found that raters inflated ratings when they were motivated to avoid a negative confrontation with poor performers, but did not adjust ratings downward when good performers rated themselves unfavorably (Shore & Tashchian, 2002).

Two other issues regarding accountability merit mention. First, a compelling piece by Walker and Smither (1999) looked at accountability from the manager's perspective (i.e., manager as a ratee) in an upward feedback context. Managers' performance improved over a 5-year period after participating in an upward feedback system, but even more interesting was the finding that improvement was more likely when managers were part of feedback sessions with their direct reports. In other words, the implication is that when held accountable to meet with their subordinates about their feedback, managers were more likely to use the feedback and improve performance.

Finally, there has also been a call from practitioners to use accountability as a means of improving the accuracy of appraisal ratings, increasing acceptance of the appraisal system, and making HR systems more efficient. For example, many companies have applied accountability pressures on managers to improve coaching as well as diversity initiatives. Companies like Motorola, Proctor & Gamble, Sara Lee, Texaco, and Steelcase have all introduced accountability pressures to further their own diversity initiatives (Digh, 1998).

Ratee issues. A second major focus of PA research since the 1990s consists of research centered on the PA ratee. Two areas, in particular, were uncovered by our literature review: the role of PA in ratee motivation and ratee reactions to PA processes. The research focusing on motivation seems to be rather easily categorized as being about either (1) the links between performance ratings and rewards or (2) those elements of the performance appraisal process which increase ratees' motivation such as participation. One theme of some recent work is that although *merit pay systems* (e.g., Pay for Performance, Performance Pay, etc.) sound like a good idea, there is very little research indicating that they are at all successful (Campbell, Campbell & Chia, 1998; Goss, 2001). Campbell et al. (1998) argue that in spite of its intuitive appeal and theoretical support, merit pay plans seldom reach their objectives. These authors suggest replacing individual-level merit pay systems with work-unit based merit pay systems that would track work group performance using performance indicators or ratios. While suggesting this alternative, the authors also recognize that some of the problems with individual-level merit pay systems may simply be elevated to the work-unit level. We believe that more empirical work is in order before this approach could be safely implemented in typical organizations.

A related article argues that while pay is an important motivator along with recognition, work enjoyment, and self-motivation, very few organizations actually link the PA system

to pay or compensation in any clear, tangible way (Mani, 2002). This same article presents the results from a survey about a university performance management system where few dollars were available to be allocated based on the performance appraisal system and those dollars that were available were distributed to all employees who were rated at least “below good.” Very little added bonus was available for being one of the best performers in the university. Finally, there has been an occasional article per year since the mid-1990s which takes Deming’s perspective that PA is one of the seven deadly sins of management (Starcher, 1996). These scholars argue that how well employees perform is much more a function of the situational constraints they experience than their own skill or motivation. Certainly, these situational constraints are important, but we would argue not so important to exclude the social or motivational factors that have been quite clearly linked to employee satisfaction and productivity over the years. We turn now to recent research that has examined the relationships between various appraisal factors and ratee motivation.

Both traditional academic research (e.g., Pettijohn, Pettijohn & d’Amico, 2001b; Roberts & Reed, 1996) and more practitioner-focused research (e.g., Pettijohn, Pettijohn, Taylor & Keillor, 2001a; Roberts, 2003; Shah & Murphy, 1995) have recently identified the significance of *participation* in the appraisal process as an antecedent of ratees’ work motivation. Roberts (2003) suggests that participation is simply essential to any fair and ethical appraisal system. In two different surveys of salespeople, Pettijohn and his colleagues (Pettijohn et al., 2001a, 2001b) identified participation and perceptions of fairness as integral to employees’ perceptions of job satisfaction and organizational commitment. They conclude from these data that PA systems can be used to actually improve employees’ levels of job satisfaction, organizational commitment, and work motivation. Finally, two recent reviews and potentially agenda-setting papers lay out models of the appraisal process that very clearly highlight both participation and justice as integral to the motivational function of a PA system (Bartol, 1999; Roberts & Reed, 1996). Bartol (1999) takes an Agency Theory perspective in developing a performance management model with a focus on compensation. She proposes that an Agency Theory-based compensation system impacts on goals, rewards, and justice perceptions that determine employee levels of satisfaction and commitment as well as performance and turnover intentions. Although not employing Agency Theory in their work, Roberts and Reed (1996) take a somewhat similar tack in proposing that participation, goals, and feedback impact on appraisal acceptance which affects appraisal satisfaction and finally employee motivation and productivity. We believe that the specific paths proposed in these models ought to be tested in both laboratory and field research. These models have widespread implications for companies at both the individual and organizational level as the links between basic-level constructs such as goals and participation could be examined and tied to employee attitudes, employer–employee relationships, employee performance, organizational effectiveness, and employee withdraw behaviors (e.g., absenteeism and turnover).

Perhaps no area within the PA literature has seen as dramatic an increase in research attention since 1990 as *ratee reactions* to PA processes. Since 1995 alone, over 20 articles were uncovered by our search for “reactions” to the appraisal process or performance feedback. This interest appears to be a direct result of the transition from a measurement-based focus on performance appraisal to a social context focus. Performance appraisals are no longer just about accuracy, but are about much more including development, ownership, input, perceptions of being valued, and being a part of an organizational team. We have

grouped these articles into three clusters: (1) reactions to the appraisal process, (2) reactions to the appraisal structure or format, and (3) reactions to multi-source appraisal or feedback. The first and second will be discussed in this section and the third in the Structural Proximal section, later in the paper.

First, with a focus on reactions to the appraisal process, [Keeping and Levy \(2000\)](#) followed from earlier work by [Cardy and Dobbins \(1994\)](#) in arguing that perhaps the best criterion to use in evaluating performance appraisal systems was the reactions of ratees. The claim was that even the most psychometrically-sound appraisal system would be ineffective if ratees (and raters) did not see it as fair, useful, valid, accurate, etc. Good psychometrics cannot make up for negative perceptions on the part of those involved in the system. Based on some of their earlier work ([Cawley, Keeping & Levy, 1998](#)), these authors conducted a study to evaluate the status of the measurement of the most common performance appraisal reactions. Their results suggested that the most established measures of system satisfaction, session satisfaction, perceived utility, perceived accuracy, procedural justice, and distributive justice all measured these constructs quite well ([Keeping & Levy, 2000](#)). They also found results that supported a higher order appraisal reactions model that fit nicely within *appraisal effectiveness* which was defined by [Cardy and Dobbins \(1994\)](#) as the multidimensional construct or ultimate criterion for measuring the success of appraisal systems. [Figure 2](#) presents our integration of the work done on appraisal reactions in recent years into the more global theorizing of Cardy and Dobbins showing that while historically research has focused on two of the three second order constructs (errors and accuracy), recent work has now begun to make progress on the third—appraisal reactions ([Figure 2](#)).

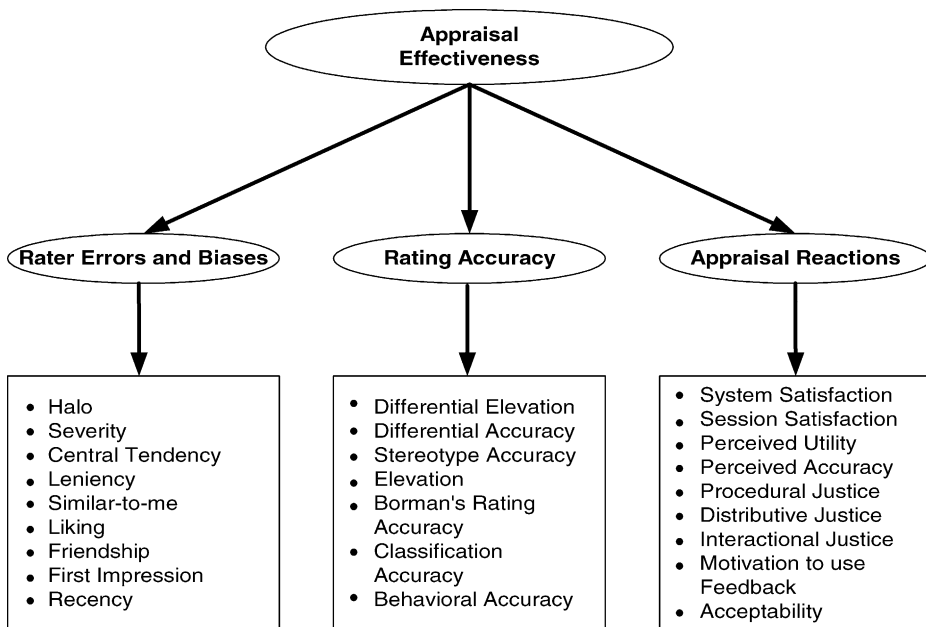


Figure 2. Appraisal effectiveness.

Many other studies have examined reactions to the appraisal process as a function of various appraisal-related variables. For instance, based on a body of work that addressed the legal approach to dispute resolution, Folger et al. (1992) applied *due process* to performance appraisal. They define three elements that must be present to achieve higher perceptions of fairness: adequate notice, fair hearing, and judgment based on evidence. Although they identified specific interventions that should be implemented to increase due process, they cautioned that, “due process mechanisms must be implemented in terms of guiding principles (i.e., designed with process goals in mind) rather than in a legalistic, mechanical, rote, or “cookbook” fashion” (p. 147).

Taylor, Masterson, Renard and Tracy (1998) conducted an initial test of this model and found that ratees appraised within a due process approach reported more positive appraisal perceptions (e.g., satisfaction with appraisal system and rating, higher perceptions of fairness and rating accuracy). More recent work conducted by Erdogan, Kraimer and Liden (2001) also supported the positive effects of due process on appraisal outcomes. Specifically, they found that elements of due process (e.g., knowledge of criteria, fair hearing) were differentially related to system and rater procedural justice perceptions.

In general, studies have found that both ratees and raters respond more favorably to fair performance appraisal systems (e.g., less emotional exhaustion, more acceptance of the feedback, more favorable reactions toward the supervisor, more favorable reactions toward the organization, and more satisfaction with the appraisal system and the job on the part of both rater and ratee) (Brown & Benson, 2003; Flint, 1999; Leung, Su & Morris, 2001; Levy & Williams, 1998; Taylor et al., 1998, 1995). Also, some research has investigated ratee appraisal reactions as a function of being provided more opportunities for participation or more information about the appraisal process (Cawley et al., 1998; Levy & Williams, 1998; Williams & Levy, 2000). In a meta-analysis of the relationship between ratee participation and reactions, Cawley et al. found strong consistent relationships between various forms of participation (e.g., value-expressive voice and instrumental voice) and the typical ratee reactions. They emphasized that even when voice was only perceived to be a way to express one’s values and not a way to affect the ensuing decision, it was still strongly related to the reactions—in fact, these relationships were somewhat stronger than they were when voice was believed to be instrumental to the decision making.

Second, a handful of studies have looked at how individuals react to elements of appraisal systems that are more structural in nature. For instance, Tziner and his colleagues (Tziner & Kopelman, 2002; Tziner, Kopelman & Joanis, 1997) have examined rater and ratee reactions to different performance appraisal formats. They found that both raters and ratees responded more favorably to behavior observation scales (BOS) than they did to other scales such as graphic rating scales or BARS. In their recent review (Tziner & Kopelman, 2002), they argue that there is some evidence for the superiority of the BOS over other rating formats with regard to rater and ratee reactions, but also that the differences between these reactions and those toward graphic rating scales are sometimes rather small. Also, their review concludes that the BARS is generally not well received by raters or ratees and consistently ranks below the other two ratings formats. DeNisi and his colleagues have focused on cognitive techniques to improve the appraisal process and, in particular, have examined the role of diary-keeping which is a procedure in which raters observe and record information about ratees in some formal (usually written) way. They have found

that raters react more favorably to PA systems that employ diaries even though in many instances it is more work for them (DeNisi & Peters, 1996). They found that raters are better able to recall performance information and were better able to discriminate among employees.

Leader–member dyadic issues. Over the course of the last decade, research on the relationships between leaders and their members has advanced considerably. Researchers have posited that trust is a key element in managing the supervisor–employee relationship (Patton, 1999). According to Mayer and Davis (1999) trust is made up of three components: ability, benevolence, and integrity. In other words, if an employee believes a supervisor has the skills to properly appraise, has the interests of the employee at the heart, and believes the supervisor upholds standards and values, the employee is likely to trust that supervisor. Interest in understanding the processes related to trust are the result of research that supports both the direct and indirect effects of trust on important organizational and individual outcomes. For instance, research has supported the relationship between trust and outcomes such as employee attitudes, cooperation, communication, and organizational citizenship behaviors (Dirks & Ferrin, 2001).

As with appraisal perceptions and reactions, researchers also believe that trust issues can limit the effectiveness of performance appraisal. For instance, if ratees have low levels of trust for their supervisor, they may be less satisfied with the appraisal and may not as readily accept feedback from that source. Hedge and Teachout (2000) examined predictors of acceptability and found that trust associated with other raters, the appraisal process, and the researchers were all significant predictors of appraisal acceptability for both job incumbents and supervisors. Similarly, Mani (2002) examined employee attitudes related to appraisal and found that trust in supervisors was important for determining satisfaction with the appraisal system.

Other researchers have examined factors that influence trust within the performance appraisal process. For instance, Whitener, Brodt, Korsgaard and Werner (1998) provide a framework that identifies possible organizational and individual factors that determine perceptions of manager's trustworthiness, along with behaviors that define trustworthy behavior. Korsgaard and Roberson (1995) found that when employees were given assertiveness training and the opportunity to self-appraise, they reported greater trust in the manager and more positive attitudes toward the appraisal. Mayer and Davis (1999) found that when a performance appraisal system was "acceptable" (e.g., perceived as accurate and being high in instrumentality), employees reported higher levels of trust for management.

The second major variable that has been considered frequently within the category of leader–member relationship is the *dyadic relationship* between the leader and employee. Leader–member exchange (LMX) theory was developed to capture the process through which leaders respond to and interact with subordinates. This aspect of the social milieu has been of interest within the larger I/O, HR, and OB disciplines (Graen & Scandura, 1987; Liden, Sparrowe & Wayne, 1997) and more specifically within the performance appraisal literature (Duarte, Goodson & Klich, 1993; Kacmar, Witt, Zivnuska & Gully, 2003; Varma & Stroh, 2001; Vecchio, 1998).

Leader–member exchange theory suggests that leaders differentially interact, respond to, and treat subordinates depending upon their membership in the "in" or "out" groups. Work

by Duarte and colleagues (Duarte & Goodson, 1994; Duarte et al., 1993) provided evidence that the relationship between objective measures of performance and supervisor ratings of performance were moderated by LMX such that in-group members were rated higher regardless of objective levels of performance. Vecchio (1998) attempted to replicate these findings, but did not find support for this effect. He suggested that perhaps those interactions were context specific and concluded that perhaps the quality of the relationship doesn't bias ratings of performance. Kacmar et al. (2003) examined whether communication frequency would moderate the relationship between LMX and supervisory ratings. They found, across two studies, that individuals who communicated with their supervisor more frequently and were in a high LMX relationship received the highest performance ratings. Interestingly, individuals with high communication frequency and in low LMX relationships received the lowest performance ratings.

Two recent studies have examined the impact of leader gender on performance ratings. Varma and Stroh (2001) proposed that same-sex leader-member dyads would result in greater liking of subordinates, which in turn would result in higher LMX relationships. Results of their study suggested that the composition of the dyad did have an effect on affect, LMX, and performance ratings. Furnham and Stringfield (2001) examined whether female and male managers rated performance differently. They found that male employees received lower ratings than female employees. In addition they found that female managers, as compared to male managers, rated male employees lower than female employees. Clearly additional research is needed to more fully examine any potential rating bias due to gender.

Group dynamics. In addition to purely dyadic issues between a supervisor and subordinate, there has been a growing concern in the performance appraisal literature about other multiple, complex relationships that impact on the appraisal process. In this section, we consider some of the research that has focused on these more general issues revolving around group, team, or workforce composition. The literature review uncovered quite a few articles that suggest at least three categories requiring our attention: (1) politics and impression management, (2) work group or team processes, and (3) the feedback environment or culture experienced by organizational employees.

In a thoughtful and integrative piece, Kozlowski and his colleagues argue that performance appraisal is a ripe situation for those involved to play political games via, among other approaches, rater distortion of ratings or ratees' active management of impressions (Kozlowski et al., 1998). Their chapter presents a case study of the political process in a military setting highlighting many of the opportunities and pressures that play into the political climate of the appraisal process. Clint Longenecker wrote a few influential papers in the 1980s where he argued that performance appraisal may be used by many raters as a political process for rewarding and punishing subordinates (Longenecker, Sims & Gioia, 1987). These authors interviewed 60 executives regarding the politics involved in performance appraisal and found great consistency of opinion among those participants that politics played a very large role in the appraisal process. However, very little empirical work has been conducted to help us understand how politics actually affects the appraisal process. Although perceptions of politics have begun to garner more attention in the I/O, HR, and OB literatures such as the work done by Ferris and his colleagues (Ferris & King, 1991, 1992), we don't have much more than some suggestions and anecdotal evidence to offer

performance appraisal practitioners. This is clearly a direction for future research as most researchers and practitioners would agree that politics has the potential to play an important role in the appraisal process.

Unlike with politics, there has been more empirical work done on the role of *impression management* in performance appraisal in recent years. One of the stronger and more sophisticated research efforts was conducted by [Wayne and Liden \(1995\)](#) in which they tested a theoretical model that proposed a series of relationships between various types of impression management such as supervisor-focused (e.g., ingratiation) or self-focused (e.g., boasting) and liking. In general, they found strong support for their model including their prediction that self-focused impression management would lead to perceptions of less similarity to the subordinate and, thus, lower ratings. In a second study along these lines, researchers found that assertive impression management techniques (e.g., ingratiation and self-promotion) resulted in higher performance ratings than did defensive impression management techniques (e.g., excuse-making and justifications) ([Gendersen & Tinsley, 1996](#)).

Another area of importance that we need to consider is performance appraisal in a team-based environment. Unfortunately, while a good bit of work has been conducted on the metrics involved in team performance measurement, very little empirical work has been conducted regarding how to develop and implement appraisals in this context despite the great influx of team-based work environments. Doing performance appraisal in a team-based environment is complicated for a few reasons. First, it is imperative that the appraisal system balances the individual vs. the team. Both are important and emphasizing individual or team performance at the exclusion of the other will result in an ineffective system. Second, it's also complicated because the PA system needs to be broad enough to include nontraditional performance criteria such as teamwork or cooperation. [Levy and Steelman \(1997\)](#) considered these complexities in proposing a prototypical appraisal system to be used in a team-based environment. This team-based model would certainly require adjustments to fit a particular organization, but the generic model includes multi-source ratings of both individual and team performance (e.g., production quality, technical knowledge, functioning of team, customer satisfaction), objective measures of individual and team performance (e.g., scrap rate, production quantity, achievement of team goals and safety objectives) as well as measures of teamwork (e.g., communication, coordination, and conflict resolution skills).

Although no empirical work exists that builds from their contextual framework, a survey of 35 organizations doing peer review across the United States found that making customers an important part of the team review was rated as very important as was getting feedback from each team member rather than just a subset of team members and the manager ([Hitchcock, 1996](#)). Rating team as well as individual performance was valued by the respondents as well—each of these points is consistent with the Levy and Steelman proposed system. In a quasi-experiment in which one class was assigned to do peer ratings and another was not, [Erez, Lepine and Elms \(2002\)](#) proposed that doing peer evaluations within the classroom setting would lead to more sharing of the group's workload, more voice in the process, and more cooperation among the group members. Further, they argued that these process variables would mediate the relationship between the conditions (i.e., peer rating vs. no peer ratings) and performance. The results were quite impressive as the relationships emerged much as the authors hypothesized suggesting that a simple peer appraisal

system has favorable effects on team member attitudes and behaviors culminating in team performance. We encourage scholars to build on the theoretical work that has begun and to make use of the various performance metrics that exist, but to take this work in a different direction—focusing on the performance appraisal process rather than solely on the more traditional measurement of team effectiveness.

The last element that we've included in this section on group dynamics is what we call the feedback environment and what others such as London call the feedback culture (London & Smither, 2002). London (2003) argues that a feedback-oriented culture is characterized by managers and employees feeling comfortable both providing and receiving feedback. Further, feedback is an integral component of the performance management process in organizations characterized by a feedback-oriented culture. London's argument is that we need to do a much better job examining the feedback culture of organizations so that we can empirically link it to other elements of the performance management cycle (e.g., employees' feedback orientation, goals, perceptions, and behavior change). Finally, the feedback culture of the organization should play a vital role in how feedback is sought, perceived, processed, accepted, used, and reacted to. In other words, the entire feedback process which is so vital to performance appraisal is, in many ways, affected by the feedback culture. We agree with London that this is an area where future research is called for and can play an instrumental role in gaining a better understanding of the performance management process.

Although there has not been a great deal of research related to this approach or linking some of the established findings of the literature to these ideas, there has been some very recent work that fits nicely with the notion of the feedback culture. First, Levy and his colleagues (Norris-Watts & Levy, *in press*; Steelman, Levy & Snell, 2004) have developed and validated a measure of the feedback environment (the FES) that diagnoses the extent to which an organization supports the feedback processes, getting at many of the important elements discussed in London's work (2003). They measure the feedback environment by focusing on the employee's perceptions of feedback source credibility, feedback quality, feedback delivery, frequency of both diagnostic favorable and unfavorable feedback, source availability, and the extent to which feedback seeking is encouraged (Steelman et al., 2004). Each of these dimensions is considered separately with respect to the supervisor and co-worker as feedback source. This scale has been empirically validated in two different companies and relationships among the various dimensions of the scale and variables like satisfaction with feedback, motivation to use feedback, and feedback seeking were as predicted by their theory. In a follow-up to this study, it was demonstrated that a more favorable feedback environment lead to higher levels of affective commitment and OCBs leading the authors to conclude that using this instrument as a diagnostic tool to identify coaching strengths and weaknesses of supervisors has great potential for improving the performance management process (Norris-Watts & Levy, *in press*).

A second line of work that also emphasizes the importance of the feedback culture has been lead by Todd Maurer and his colleagues (Maurer, Mitchell & Barbeite, 2002; Maurer, Weiss & Barbeite, 2003). They have developed and tested a model that gives arguably the best snapshot of relationships among key constructs involved in employee learning and development (Maurer et al., 2003). Among the key antecedents to successful learning and development in their model are learning preparedness, situational support for development, and self-efficacy for development. They argue that both work and nonwork situations are

supportive of learning and development which may, of course, impact perceptions of potential benefits, attitudes, and intentions to participate in the developmental process. The way in which supportive work situations are linked to other constructs in the model is quite consistent with the work described above in which others (London, 2003; Norris-Watts & Levy, in press; Steelman et al., 2004) suggest that the dynamics of the feedback culture or environment are so integral to performance management as well as coaching and development.

Structural Proximal Variables

As noted in Figure 1, structural process variables are factors that have direct effects on rater and ratee behavior and are directly affected by distal variables. Structural variables are those aspects of the system that make up the organization or design of the performance management process. For instance, the types and number of performance dimensions that are rated, the frequency of the appraisal, and the purpose of the appraisal are all aspects of performance appraisal structure. Probably the greatest structural change that has occurred over the last 10–15 years is the implementation of multi-source (i.e., 360-degree) feedback systems. Research relevant to 360-degree feedback as well as research related to other structural issues (e.g., performance appraisal purpose and rater training) is reviewed below.

Multi-source feedback systems. Multi-source feedback systems have been implemented in organizations largely as a means to provide developmental feedback for employees (Garavan, Morley & Flynn, 1997). The benefits of a multi-source vs. traditional feedback system are predicated on three important assumptions (Borman, 1997): (1) that each of the rating sources can provide unique information about the target, (2) that these multiple ratings will provide incremental validity over individual sources, and (3) that feedback from multiple sources will increase the target's self-awareness and lead to behavioral change (Fletcher & Baldry, 2000).

Consistent with the historical review of performance appraisal literature laid out in the introduction of this paper, the early 360-degree literature has primarily been conducted from a test metaphor. In other words, the majority of the empirical work has focused on the psychometric properties of multi-source ratings (Conway & Huffcutt, 1997; Scullen, Mount & Judge, 2003), the correlations between ratings sources (Atkins & Wood, 2002; Beehr, Ivanitskaya, Hansen, Erofeev & Gudanowski, 2001; Conway & Huffcutt, 1997; Fecteau & Craig, 2001; Warr & Bourne, 1999), rater and ratee effects on ratings (Antonioni & Park, 2001; Brutus, Fleenor & McCauley, 1999; Fletcher & Baldry, 2000; Warech & Smither, 1998; Warr & Bourne, 1999) and the application of various data analytic techniques to multi-source ratings (Barr & Raju, 2003; Fecteau & Craig, 2001; Mount & Judge, 1998; Penny, 2003; Yammarino, 2003). Although this approach is important and has proved useful, we contend that for organizations to reap the full benefits of these systems, research examining the social context within which the multi-source feedback takes place will be critical. Certainly some research has begun to examine these types of issues (Atwater, Roush & Fischthal, 1995; Bernardin, Dahmus & Redmon, 1993; Hazucha, Hezlett & Schneider, 1993), but much more research is needed.

For instance, a fair amount of research has begun to examine participants' reactions to the multi-source feedback process (Albright & Levy, 1995; Brett & Atwater, 2001; Funderburg & Levy, 1997; Levy, Cawley & Foti, 1998; Williams & Lueke, 1999). For instance, Waldman and Bowen (1998) examined both rater and ratee acceptability of the process and proposed that factors such as organizational culture, credibility of raters, high rates of participation, and anonymity of ratings may be likely to influence acceptance of multi-source feedback. Williams and Lueke (1999) found that knowledge of and experience with the multi-source system as well as social support played an important role in multi-source system reactions, perceived developmental constraints and self-efficacy judgments related to development, which were important predictors of managers intentions to develop. Mauer et al. (2002) examined the effect of 360-degree ratings and individual and organizational characteristics on system attitudes and developmental activity. They found that a supportive work context and an individual's self-efficacy were the most important predictors of multi-source feedback attitudes and frequency of developmental activities. Finally, Brett and Atwater (2001) examined the relationships between rating discrepancies, feedback reactions, and receptivity to development. Their study suggests that when individuals receive lower than expected ratings, they respond quite negatively and this may influence their developmental responses. The work on multi-source feedback reactions supports the notion that elements of the social context are critical factors impacting the success of multi-source feedback systems.

Kluger and Denisi (1996) conducted an important meta-analysis on the effect of feedback interventions on performance improvement and found that feedback interventions are not as uniformly successful as we might have believed. These equivocal results seem to generalize to the multi-source feedback literature as well. Specifically, Seifert, Yuki and McDonald (2003) reviewed 14 studies that included either upward or 360-degree feedback and found that while some reported performance improvements (Atwater et al., 1995; Walker & Smither, 1999), some did not (Atwater, Waldman, Atwater & Cartier, 2000; Johnson & Ferstl, 1999), and others reported inconclusive results (Reilly, Smither & Vasilopoulos, 1996; Smither, London, Vasilopoulos, Reilly, Millsap and Salvemini, 1995).

The variability in these results suggests that perhaps other factors play a part in whether feedback results in actual performance improvement. For instance, London, Smither and Adsit (1997) discuss the role of accountability in the multi-source feedback process. Smither, London, Flautt, Varagas and Kucine (2003) examined whether the use of an executive coach following multi-source feedback resulted in greater behavioral change. They found that individuals who used an executive coach did, in fact, set more specific goals, sought out others for information, and had slightly higher ratings in subsequent multi-source appraisals. Another factor that may play a significant role in whether or not individuals actually use 360-degree feedback is participants' attitudes and reactions toward the feedback and the appraisal system (Ilgen, Fisher and Taylor, 1979). Researchers have suggested that if participants do not perceive the system to be fair, the feedback to be accurate, or sources to be credible then they are more likely to ignore and not use the feedback they receive (Fecteau & Fecteau, 1998; Ilgen et al., 1979; Waldman & Bowen, 1998).

Performance appraisal purpose. A continuing area of interest in the performance appraisal literature has been what researchers have called the "performance appraisal purpose

effect” (Jawahar & Williams, 1997). In their seminal work on rating leniency, Taylor and Wherry (1951) proposed that ratings collected for administrative purposes would be more lenient than ratings collected for research or developmental purposes. Over the last few years, not much new empirical work has been conducted (Greguras, Robie, Schleicher & Goff, 2003; Harris, Smith & Champagne, 1995), however, a useful empirical review of this literature has been written (Jawahar & Williams, 1997). Jawahar and Williams (1997) examined data from 22 studies and found that ratings collected for administrative purposes were in fact more lenient than ratings collected for research or developmental purposes. Moreover, they identified significant moderators of this relationship (e.g., study setting, type of rater, direction of feedback, and type of rating scale). While the majority of the research on performance appraisal purpose has focused on the rater, some work has also been conducted on ratee effects (Boswell & Boudreau, 2000, 2002). Given the importance of ratee reactions for the success of appraisal systems, this work is important and future research investigating these effects in other contexts (e.g., multi-source feedback systems) would be helpful.

Rater training. A great deal of performance appraisal research conducted in the 1980s focused on the effects of rater training (Bernardin & Buckely, 1981; Pulakos, 1984). A meta-analysis conducted during this time provided support for one type of training in particular, frame-of reference (FOR) training (Woehr and Huffcutt, 1994). Subsequent research on rater training over the last 10 years has focused almost primarily on FOR-training (Day & Sulsky, 1995; Keown-Gerrard & Sulsky, 2001; Noonan & Sulsky, 2001; Schleicher, Day, Mayes & Riggio, 2002; Sulsky & Day, 1992; Sulsky & Keown, 1998; Sulsky, Skarlicki & Keown, 2002). Issues recently examined are the use of FOR training to improve assessment center ratings (Schleicher et al., 2002), a comparison of FOR and behavioral observation training on rating accuracy (Noonan & Sulsky, 2001), and the effects of instructional interventions on FOR effects (Keown-Gerrard & Sulsky, 2001). The recent research on performance appraisal training appears to focus on ways to fine tune this training and apply it within other contexts. Overall, this work provides continued support for the efficacy of FOR-training.

Conclusion

Our goal in completing this review was to examine the extent to which researchers have heeded Bretz et al.’s (1992) call to better understand the social context of performance appraisal. For example, research in the last few years on the feedback culture or environment has suggested completely new approaches to performance management and coaching that were not previously well established or even considered. In fact, it seems to us that we now have well-developed theoretical frameworks, measurement technologies, and some early empirical results suggesting that the dynamic nature of the feedback environment is important. A second area that has emerged as extremely important in the recent PA literature is the newer ways in which appraisal systems are evaluated. Our review of the literature and its placement in the historical context in which PA has developed (Farr & Levy, *in press*) resulted in our model of Appraisal Effectiveness (Figure 2). We think this model accurately portrays

the ways in which the effectiveness or success of performance appraisal systems can be evaluated. Our review indicates that Appraisal Reactions is where there has been the most growth in the PA research since 1995 and also where practitioners see the most potential benefit.

Third, it is clear that more empirical work should be conducted to better isolate and understand the various relationships discussed throughout the paper. This observation is exciting and we are hopeful that the current review will help identify and clarify new research avenues for researchers. In writing this paper we observed, that often, research areas tended to focus on either rater or ratee effects, but often neglected to examine the effects of variables simultaneously on both participants. In some instances, this singular focus makes sense (e.g., rater training). However, in other areas, this focus on either the rater or ratee seems to leave the other side of the coin unexamined. Although much has been learned in various studies focused on either rater or ratee variables, we believe that understanding the PA process would be well-served by research examining both sides of the coin simultaneously.

Finally, while it appears that these initial studies are yielding useful information, it will take time to see whether these results actually benefit the practice of appraisal. Our review suggests that as a field, we seem to be moving in that direction, however, the goal should continue to be two-pronged: (1) gain a better understanding of the PA process and (2) apply that enhanced understanding to organizations so as to improve performance appraisals in use. The focus on the social context of PA has taken us down the appropriate road, but there are still many more miles to cover.

Acknowledgment

We thank Chris Rosen, Samantha Chau, and Brian Whitaker for their help in searching the literature and organizing the database.

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(7): 577.
- Allen, T. D., & Rush, M. C. 1998. The effects of organizational citizenship behavior on performance judgments: A field study and a laboratory experiment. *Journal of Applied Psychology*, 83(2): 247–260.
- Antonioni, D., & Park, H. 2001. The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management*, 27(4): 479–495.
- 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(4): 871–904.
- Atwater, L., Rousch, P., & Fischthal, A. 1995. The influence of upward feedback on self and follower ratings of leadership. *Personnel Psychology*, 48: 35–59.
- Atwater, L., Waldman, D., Atwater, D., & Cartier, P. 2000. An upward feedback field experiment: Supervisors' cynicism, reactions, and commitment to subordinates. *Personnel Psychology*, 53: 275–297.
- Bannister, B. D. 1986. Performance outcome feedback and attributional feedback: Interactive effects on recipient responses. *Journal of Applied Psychology*, 71: 203–210.
- Barr, M. A., & Raju, N. S. 2003. IRT-based assessments of rater effects in multiple-source feedback instruments. *Organizational Research Methods*, 6(1): 15–43.

- Bartol, K. M. 1999. Reframing salesforce compensation systems: An agency theory-based performance management perspective. *Journal of Personal Selling & Sales Management*, 19(3): 1.
- Bates, R. 2002. Liking and similarity as predictors of multi-source ratings. *Personnel Review*, 31(5): 540–552.
- 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(7): 775.
- Bernardin, H. J., & Buckely, M. R. 1981. Strategies in rater training. *Academy of Management Review*, 6: 205–212.
- Bernardin, H. J., Cooke, D. K., & Villanova, P. 2000. Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology*, 85(2): 232–236.
- Bernardin, H. J., Dahmus, S. A., & Redmon, G. 1993. Attitudes of first-line supervisors toward subordinate appraisals. *Human Resource Management*, 32: 315–324.
- Borman, W. C. 1997. 360 degree ratings: An analysis of assumptions and a research agenda for evaluating their validity. *Human Resource Management Review*, 7(3): 299.
- Boswell, W. R., & Boudreau, J. W. 2000. Employee satisfaction with performance appraisals and appraisers: The role of perceived appraisal use. *Human Resource Development Quarterly*, 11(3): 283.
- Boswell, W. R., & Boudreau, J. W. 2002. Separating the developmental and evaluative performance appraisal uses. *Journal of Business & Psychology*, 16(3): 391–412.
- Brett, J. F., & Atwater, L. E. 2001. 360 degrees feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, 86(5): 930.
- Bretz, R. D., Milkovich, G. T., & Read, W. 1992. The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management*, 18(2): 321.
- Brown, M., & Benson, J. 2003. Rated to exhaustion? Reactions to performance appraisal processes. *Industrial Relations Journal*, 34(1): 67.
- Brutus, S., Fleenor, J. W., & McCauley, C. D. 1999. Demographic and personality predictors of congruence in multi-source ratings. *Journal of Management Development*, 18(5): 417–435.
- Campbell, D. J., Campbell, K. M., & Chia, H.-B. 1998. Merit pay, performance appraisal, and individual motivation: An analysis and alternative. *Human Resource Management*, 37(2): 131.
- Cardy, R. L., & Dobbins, G. H. 1994. *Performance appraisal: Alternative perspectives*. Cincinnati, OH: South-Western Publishing.
- 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(4): 615–633.
- Conway, J. M., & Huffcutt, A. I. 1997. Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, Peer, and self-ratings. *Human Performance*, 10(4): 331.
- Day, D. V., & Sulsky, L. M. 1995. Effects of frame-of-reference training and information configuration on memory organization and rating accuracy. *Journal of Applied Psychology*, 80(1): 158–167.
- DeNisi, A. S., & Peters, L. H. 1996. Organization of information in memory and the performance appraisal process: Evidence from the field. *Journal of Applied Psychology*, 81(6): 717–737.
- Digh, P. 1998. The next challenge: Holding people accountable. *HR Magazine*: Vol. 43. 63. Society for Human Resource Management.
- Dirks, K. T., & Ferrin, D. L. 2001. The role of trust in organizational settings. *Organization Science*, 12(4): 450–467.
- Duarte, N. T., & Goodson, J. R. 1994. Effects of dyadic quality and duration on performance appraisal. *Academy of Management Journal*, 37(3): 499.
- Duarte, N. T., Goodson, J. R., & Klich, N. R. 1993. How do I like thee? Let me appraise the ways. *Journal of Organizational Behavior*, 14(3): 239.
- 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, A., Lepine, J. A., & Elms, H. 2002. Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. *Personnel Psychology*, 55(4): 929–948.
- Facteau, C. L., & Facteau, J. D. 1998. Reactions of leaders to 360-degree feedback from subordinates and peers. *Leadership Quarterly*, 9(4): 427.
- Facteau, J. D., & Craig, S. B. 2001. Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology*, 86(2): 215–227.

- Farr, J. L., & Levy, P. E. (in press). Performance appraisal. In L. L. Koppes (Ed.), *The science and practice of industrial-organizational psychology: The first hundred years*. Erlbaum.
- Feldman, J. M. 1981. Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66: 127–148.
- Ferris, G. R., Judge, T. A., Rowland, K. M., & Fitzgibbons, D. E. 1994. Subordinate influence and the performance evaluation process: Test of a model. *Organizational Behavior & Human Decision Processes*, 58(1): 101.
- Ferris, G. R., & King, T. R. 1991. Politics in human resources decisions: A walk on the dark side. *Organizational Dynamics*, 20(2): 59.
- Ferris, G. R., & King, T. R. 1992. The politics of age discrimination in organizations. *Journal of Business Ethics*, 11(5/6): 341.
- Fletcher, C. 2001. Performance appraisal and management: The developing research agenda. *Journal of Occupational & Organizational Psychology*, 74(4): 473–487.
- Fletcher, C., & Baldry, C. 2000. A study of individual differences and self-awareness in the context of multi-source feedback. *Journal of Occupational & Organizational Psychology*, 73(3): 303–319.
- Flint, D. H. 1999. The role of organizational justice in multi-source performance appraisal: Theory-based applications and directions for research. *Human Resource Management Review*, 9(1): 1.
- Folger, R., Konovsky, M., & Cropanzano, R. 1992. A due process metaphor for performance appraisal. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior*: Vol. 14. 129–177. Greenwich, CT: JAI.
- Forgas, J. P., & George, J. M. 2001. Affective influences on judgments and behavior in organizations: An information processing perspective. *Organizational Behavior and Human Decision Processes*, 86(1): 3–34.
- Fried, Y., Tieg, R. B., & Bellamy, A. R. 1992. Personal and interpersonal predictors of supervisors' avoidance of evaluating subordinates. *Journal of Applied Psychology*, 77(4): 462–468.
- Frink, D. D., & Ferris, G. R. 1998. Accountability, impression management, and goal setting in the performance evaluation process. *Human Relations*, 51(10): 1259.
- Funderburg, S. A., & Levy, P. E. 1997. The influence of individual and contextual variables on 360-degree feedback system attitudes. *Group & Organization Management*, 22(2): 210.
- Furnham, A., & Stringfield, P. 2001. Gender differences in rating reports: Female managers are harsher raters, particularly of males. *Journal of Managerial Psychology*, 16(4): 281.
- Garavan, T. N., Morley, M., & Flynn, M. 1997. 360 degree feedback: Its role in employee development. *Journal of Management Development*, 16(2/3): 134.
- Gendersen, D. E., & Tinsley, D. B. 1996. Empirical assessment of impression management biases: The potential for performance appraisal error. *Journal of Social Behavior and Personality*, 11(5): 57–77.
- Goss, W. 2001. Managing for results—Appraisals and rewards. *Australian Journal of Public Administration*, 60(1): 3.
- Graen, G. B., & Scandura, T. A. 1987. Toward a psychology of dyadic organizing. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior*: Vol. 9. 175–208. Greenwich, CT: JAI.
- Greguras, G. J., Robie, C., Schleicher, D. J., & Goff, M. 2003. A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology*, 56(1): 1–21.
- Harris, M. M. 1994. Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management*, 20(4): 737.
- Harris, M. M., Smith, D. E., & Champagne, D. 1995. A field study of performance appraisal purpose: Research versus administrative-based ratings. *Personnel Psychology*, 48(1): 151.
- Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. 1993. The impact of 360-degree feedback on management skills development. *Human Resource Management*, 32: 251–325.
- Hebert, B. G., & Vorauer, J. D. 2003. Seeing through the screen: Is evaluative feedback communicated more effectively in face-to-face or computer-mediated exchanges? *Computers in Human Behavior*, 19(1): 25–38.
- Hedge, J. W., & Teachout, M. S. 2000. Exploring the concept of acceptability as a criterion for evaluating performance measures. *Group & Organization Management*, 25(1): 22–44.
- Hitchcock, D. 1996. What are people doing around peer review? *Journal for Quality & Participation*, 19(7): 52.
- Ilgen, D. R., Barnes-Farrell, J. L., & McKellin, D. B. 1993. Performance appraisal process research in the 1980s: What has it contributed to appraisals in use? *Organizational Behavior and Human Decision Processes*, 54: 321–368.

- Ilgén, D. R., Fisher, C. D., & Taylor, M. S. 1979. Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64: 349–371.
- Jawahar, I. M., & Williams, C. R. 1997. Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, 50(4): 905.
- 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(4): 808–815.
- Johnson, J., & Ferstl, K. L. 1999. The effects of interrater and self-other agreement on performance improvement following upward feedback. *Personnel Psychology*, 52: 271–303.
- Kacmar, K. M., Witt, L. A., Zivnuska, S., & Gully, S. M. 2003. The interactive effect of leader–member exchange and communication frequency on performance ratings. *Journal of Applied Psychology*, 88(4): 764–772.
- Keeping, L. M., & Levy, P. E. 2000. Performance appraisal reactions: Measurement, modeling, and method bias. *Journal of Applied Psychology*, 85(5): 708–723.
- Keown-Gerrard, J. L., & Sulsky, L. M. 2001. The effects of task information training and frame-of-reference training with situational constraints on rating accuracy. *Human Performance*, 14(4): 305.
- Klimoski, R., & Inks, L. 1990. Accountability forces in performance appraisal. *Organizational Behavior & Human Decision Processes*, 45(2): 194.
- 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.
- Korsgaard, M. A., & Roberson, L. 1995. Procedural justice in performance evaluation: The role of instrumental and non-instrumental voice in performance appraisal decisions. *Journal of Management*, 21: 657–669.
- Kozlowski, S. W. J., Chao, G. T., & Morrison, R. F. 1998. Games raters play: Politics, strategies, and impression management in performance appraisal. In J. W. Smithers (Ed.), *Performance appraisal: State of the art in practice*: 163–205. San Francisco: Jossey-Bass.
- Landy, F. J., & Farr, J. L. 1980. Performance rating. *Psychological Bulletin*, 87: 72–107.
- Lefkowitz, J. 2000. The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational & Organizational Psychology*, 73(1): 67–85.
- Leung, K., Su, S., & Morris, M. W. 2001. When is criticism not constructive? The roles of fairness perceptions and dispositional attributions in employee acceptance of critical supervisory feedback. *Human Relations*, 54(9): 1155–1187.
- Levy, P. E., Cawley, B. D., & Foti, R. J. 1998. Reactions to appraisal discrepancies: Performance ratings and attributions. *Journal of Business & Psychology*, 12(4): 437.
- Levy, P. E., & Steelman, L. A. 1997. Performance appraisal for team-based organizations: A prototypical multiple rater system. In M. Beyerlein, D. Johnson, & S. Beyerlein (Eds.), *Advances in interdisciplinary studies of work teams: Team implementation issues*: Vol. 4. 141–165. Greenwich, CT: JAI.
- 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(1): 53–65.
- Liden, R. C., Sparrowe, R. T., & Wayne, S. J. 1997. Leader–member exchange theory: The past and potential for the future. In G. R. Ferris (Ed.), *Research in personnel and human resources management*: Vol. 15. 47–119. Greenwich, CT: JAI.
- London, M. 2003. *Job feedback: Giving, seeking and using feedback for performance improvement* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- London, M., & Smithers, J. W. 2002. Feedback orientation, feedback culture, and the longitudinal performance management process. *Human Resource Management Review*, 12(1): 81.
- London, M., Smithers, J. W., & Adsit, D. J. 1997. Accountability: The achilles' heel of multisource feedback. *Group and Organization Management*, 22: 162–184.
- Longenecker, C. O., Sims, H. P., & Gioia, D. A. 1987. Behind the mask: The politics of employee appraisal. *Academy of Management Executive*, 1: 183–193.
- Mani, B. G. 2002. Performance appraisal systems, productivity, and motivation: A case study. *Public Personnel Management*, 31(2): 141–159.
- Maurer, T. J., Mitchell, D. R. D., & Barbeite, F. G. 2002. Predictors of attitudes toward a 360-degree feedback system and involvement in post-feedback management development activity. *Journal of Occupational & Organizational Psychology*, 75(1): 87–107.

- Maurer, T. J., Weiss, E. W., & Barbeite, F. G. 2003. A model of involvement in work-related learning and development activity: The effects of individual, situational, motivational, and age variables. *Journal of Applied Psychology*, 88(4): 707–724.
- 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(1): 123–136.
- Mero, N. P., & Motowidlo, S. J. 1995. Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology*, 80(4): 517–524.
- Mero, N. P., Motowidlo, S. J., & Anna, A. L. 2003. Effects of accountability on rating behavior and rater accuracy. *Journal of Applied Social Psychology*, 33(12): 2493–2514.
- Miller, J. S. 2003. High tech and high performance: Managing appraisal in the information age. *Journal of Labor Research*, 24(3): 409.
- Mount, M. K., & Judge, T. A. 1998. Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, 51(3): 557.
- Murphy, K. R., & Cleveland, J. N. 1991. *Performance appraisal: An organizational perspective*. Boston: Allyn & Bacon.
- Murphy, K. R., & Cleveland, J. N. 1995. *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks: Sage.
- Noonan, L. E., & Sulsky, L. M. 2001. Impact of frame-of-reference and behavioral observation training on alternative training effectiveness criteria in a Canadian military sample. *Human Performance*, 14(1): 3–26.
- Norris-Watts, C., & Levy, P. E. (in press). The mediating role of affective commitment in the relation of the feedback environment to work outcomes. *Journal of Vocational Behavior*.
- Patton, F. 1999. Oops, the future is past and we almost missed it!"—Integrating quality and behavioral management methodologies. *Journal of Workplace Learning*, 11(7): 266–277.
- Penny, J. A. 2003. Exploring differential item functioning in a 360-degree assessment: Rater source and method of delivery. *Organizational Research Methods*, 6(1): 61–79.
- Pettijohn, C., Pettijohn, L. S., Taylor, A. J., & Keillor, B. D. 2001. Are performance appraisals a bureaucratic exercise or can they be used to enhance sales-force satisfaction and commitment? *Psychology & Marketing*, 18(4): 337–364.
- 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(2): 127–146.
- Pulakos, E. D. 1984. A comparison of training programs: Error training and accuracy training. *Journal of Applied Psychology*, 69: 581–588.
- Reilly, R. R., Smither, J. W., & Vasilopoulos, N. 1996. A longitudinal study of upward feedback. *Personnel Psychology*, 49: 599–612.
- Robbins, T. L., & DeNisi, A. S. 1998. Mood vs. interpersonal affect: Identifying process and rating distortions in performance appraisal. *Journal of Business & Psychology*, 12(3): 313.
- Roberts, G. E. 2003. Employee performance appraisal system participation: A technique that works. *Public Personnel Management*, 32(1): 89.
- Roberts, G. E., & Reed, T. 1996. Performance appraisal participation, goal setting and feedback. *Review of Public Personnel Administration*, 16(4): 29.
- Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. 2002. A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology*, 87(4): 735–746.
- Scullen, S. E., Mount, M. K., & Judge, T. A. 2003. Evidence of the construct validity of developmental ratings of managerial performance. *Journal of Applied Psychology*, 88(1): 50–66.
- Seifert, C. F., Yukl, G., & McDonald, R. A. 2003. Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *Journal of Applied Psychology*, 88(3): 561–569.
- Shah, J. B., & Murphy, J. 1995. Performance appraisals for improved productivity. *Journal of Management in Engineering*, 11(2): 26.
- Shore, T. H., & Tashchian, A. 2002. Accountability forces in performance appraisal: Effects of self-appraisal information, normative information, and task performance. *Journal of Business & Psychology*, 17(2): 261–274.
- Sinclair, R. C. 1988. Mood, categorization, breadth, and performance appraisal: The effects of order of information acquisition and affective state on halo, accuracy, information retrieval, and evaluations. *Organizational Behavior and Human Decision Processes*, 42: 22–46.

- 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(1): 23–44.
- Smither, J. W., London, M., Vasilopoulos, N., 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.
- Starcher, R. 1996. Individual performance appraisal systems. *Production & Inventory Management Journal*, 37(4): 58.
- Steelman, L. A., Levy, P. E., & Snell, A. F. 2004. The feedback environment scale (FES): Construct definition, measurement, and validation. *Educational and Psychological Measurement*, 64(1): 165–184.
- Strauss, J. P., Barrick, M. R., & Connerley, M. L. 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 & Organizational Psychology*, 74(5): 637–657.
- Struthers, C. W., Weiner, B., & Allred, K. 1998. Effects of causal attributions on personnel decisions: A social motivation perspective. *Basic & Applied Social Psychology*, 20(2): 155–166.
- Sulsky, L. M., & Day, D. V. 1992. Frame-of-reference training and cognitive categorization: An empirical investigation of rater memory issues. *Journal of Applied Psychology*, 77(4): 501–510.
- Sulsky, L. M., & Keown, J. L. 1998. Performance appraisal in the changing world of work: Implications for the meaning and measurement of work performance. *Special Issue: Industrial-Organizational Psychology and Emerging Needs of the Canadian Workplace: Traversing the Next Millennium*, 39(1/2): 52–59.
- Sulsky, L. M., Skarlicki, D. P., & Keown, J. L. 2002. Frame-of-reference training: Overcoming the effects of organizational citizenship behavior on performance rating accuracy. *Journal of Applied Social Psychology*, 32(6): 1224–1240.
- Taylor, M. S., Masterson, S. S., Renard, M. K., & Tracy, K. B. 1998. Managers' reactions to procedurally just performance management systems. *Academy of Management Journal*, 41(5): 568–579.
- 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(3): 495–523.
- Tziner, A., & Wherry, R. J. 1951. A study of leniency in two rating systems. *Personnel Psychology*, 4: 39–47.
- Tziner, A., Kopelman, R., & Joanis, C. 1997. Investigation of raters' and ratees' reactions to three methods of performance appraisal: BOS, BARS, and GRS. *Canadian Journal of Administrative Sciences*, 14(4): 396.
- Tziner, A., & Kopelman, R. E. 2002. Is there a preferred performance rating format? A non-psychometric perspective. *Applied Psychology*, 51(3): 479.
- Varma, A., Denisi, A. S., & Peters, L. H. 1996. Interpersonal affect and performance appraisal: A field study. *Personnel Psychology*, 49(2): 341.
- Varma, A., & Stroh, L. K. 2001. The impact of same-sex LMX dyads on performance evaluations. *Human Resource Management*, 40(4): 309.
- Vecchio, R. P. 1998. Leader–member exchange, objective performance, employment duration, and supervisor ratings: Testing for moderation and mediation. *Journal of Business & Psychology*, 12(3): 327.
- Villanova, P., Bernardin, H. J., Dahmus, S. A., & Sims, R. L. 1993. Rater leniency and performance appraisal discomfort. *Educational & Psychological Measurement*, 53(3): 789–799.
- Waldman, D. A., & Bowen, D. E. 1998. The acceptability of 360 degree appraisals: A customer-supplier relationship perspective. *Human Resource Management*, 37(2): 117.
- Walker, A. G., & Smither, J. W. 1999. A five-year study of upward feedback: What managers do with their results matters. *Personnel Psychology*, 52(2): 393–423.
- Warech, M. A., & Smither, J. W. 1998. Self-monitoring and 360-degree ratings. *Leadership Quarterly*, 9(4): 449.
- Warr, P., & Bourne, A. 1999. Factors influencing two types of congruence in multirater judgments. *Human Performance*, 12(3/4): 183–210.
- Wayne, S. J., & Liden, R. C. 1995. Effects of impression management on performance ratings: A longitudinal study. *Academy of Management Journal*, 38(1): 232–260.
- Whitener, E. M., Brodt, S. E., Korsgaard, M. A., & Werner, J. M. 1998. Managers as initiators of trust: An exchange relationship framework for understanding managerial trustworthy behavior. *The Academy of Management*, 23: 531–546.
- Williams, J. R., & Levy, P. E. 2000. Investigating some neglected criteria: The influence of organizational level and perceived system knowledge on appraisal reactions. *Journal of Business & Psychology*, 14(3): 501–513.

- Williams, J. R., & Lueke, S. B. 1999. 360 degrees feedback system effectiveness: Test of a model in a field setting. *Journal of Quality Management*, 4(1): 23.
- Woehr, D. J., & Huffcutt, A. I. 1994. Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67: 189–205.
- Yammarino, F. J. 2003. Modern data analytic techniques for multisource feedback. *Organizational Research Methods*, 6(1): 6.

Paul E. Levy is a Professor of Psychology at The University of Akron where he also serves as Associate Department Chair and Chair of the I/O Psychology Program. He publishes extensively in the top journals in the field such as *Journal of Applied Psychology*, *Organizational Behavior and Human Decision Processes*, *Personnel Psychology*, *Journal of Management*, and *Journal of Personality and Social Psychology*. He is the author of a recent textbook on I/O Psychology, serves on multiple editorial boards, and regularly consults to industry.

Jane R. Williams is an Associate Professor of Psychology at Indiana University-Purdue University Indianapolis. She received her Ph.D. in Industrial/Organizational Psychology from the University of Akron in 1995. She has published her work in such journals at *Journal of Applied Psychology*, *Personnel Psychology*, *Organizational Behavior and Human Decision Processes*, and *Journal of Organizational Behavior*.