CAHRS
Center for Advanced Human Resource Studies

CAHRS / Cornell University
187 Ives Hall
Ithaca, NY 14853-3901 USA
Tel. 607 255-9358
www.ilr.cornell.edu/CAHRS/

WORKING PAPER SERIES

Separating the Developmental and Evaluative Performance Appraisal Uses

Wendy R. Boswell
John W. Boudreau

Working Paper 99–09

ilr School of Industrial and Labor Relations
CORNELL UNIVERSITY
Advancing the World of Work

Separating the Developmental

and Evaluative Performance Appraisal Uses

Wendy R. Boswell
Center for Advanced Human Resource Studies
School of Industrial & Labor Relations
Cornell University

John W. Boudreau
Center for Advanced Human Resource Studies
School of Industrial and Labor Relations
Cornell University

Working Paper 99-09

http://www.ilr.cornell.edu/cahrs

Correspondence concerning this paper should be addressed to Wendy R. Boswell, Department of Human Resource Studies, School of Industrial and Labor Relations, 393 Ives Hall, Cornell University, Ithaca, NY 14853-3901. Telephone: (607) 255-6552; Fax: (607) 255-1836. Electronic mail may be sent via Internet to wrh4@cornell.edu.

Separating the Developmental and Evaluative Performance Appraisal Uses

The multiple uses of performance appraisal (PA) have been the focus of much research, often yielding conflicting findings and conclusions. This study used an untreated control group design to examine the effects of separating the developmental and evaluative PA uses (in time and by person) on employee attitudes and behavioral intentions. Results indicated no effect on employee attitudes, however, employees in the separated PA group reported they were less likely to use development in the future.

Performance assessment is often considered one of the most important Human Resource (HR) practices (e.g.; Judge & Ferris, 1993). Although the need for some type of performance appraisal (PA) system is usually acknowledged, the practical usefulness of PAs is often questioned (McNerney, 1995; Murphy & Cleveland, 1995). This is consistent with other management practices. Discussing the dynamics of bureaucracy, Blau states, “Many social patterns that served important functions for operations also had some dysfunctions, that is, they produced conditions that impeded the effective attainment of organizational objectives” (1963: 252). The PA process may be one such function, believed to be a necessity in terms of making administrative decisions, providing legal safeguards, and even improving performance (Murphy & Cleveland, 1995), yet at risk of promoting negative outcomes particularly if employed for multiple uses (Deming, 1986; DeNisi & Williams, 1988; Meyer, Kay, & French, 1965).

Previous research and theory indicate that there is a conflict when PA is used both to develop and evaluate (Bernardin & Beatty, 1984; DeNisi, Cafferty, & Meglino, 1984; Longenecker, Gioia, & Sims, 1987; Murphy, Balzer, Kellam, & Armstrong, 1984; Murphy & Cleveland, 1995; Williams, DeNisi, Blencoe, & Cafferty, 1985; Zedeck & Cascio, 1982). Practitioner articles have also advocated separating evaluation (e.g., salary administration) from employee development, even suggesting that employees cannot receive accurate feedback about their strengths and developmental needs if these two PA uses are connected (e.g., Harvey, 1995; McNerney, 1995). This conflict is believed to prevent the appraisal process from attaining its goals and usefulness to the organization, perhaps even contributing negatively to individual behaviors and organizational performance. Yet, research actually examining this conflict is rare. We know little about whether employees react or behave differently when evaluation and development are combined within one PA process versus when they are separated. The present study begins to address this issue by investigating the effects of separating PA uses on employee attitudes and behavioral intentions.

The present research views PA not just as a single measurement event, but instead as a process. This perspective follows Murphy and Cleveland’s (1995) view of PA as a means for

social interaction and communication. Their social-psychological perspective is useful in focusing PA research beyond the mechanics and measurement problems of the appraisal event, and instead toward improvements in the overall process. This perspective makes salient the issue of possible conflicting PA uses by recognizing that the appraisal process is part of a larger organizational system where organizational practices and employee experiences, although not necessarily part of the appraisal event, may still exert an influence. In other words, it is important to consider the general relationship between evaluation and development within an organizational system and its role in affecting employees' experience and reactions. In sum, by looking at PA as a process by which organizations communicate and make decisions, social and psychological issues associated with the PA can be better understood and improved.

Performance Appraisal Uses

We examine the effects of separating the evaluative and developmental components of PA, so it is important to define development and evaluation. Development is any effort concerned with enriching attitudes, experiences, and skills which improve the effectiveness of employees. Specific examples of developmental PA use include: identifying an individual's strengths and weaknesses, setting goals, and identification of training needs. Evaluation is characterized as comparing an individual's performance to a set standard, other organizational members, or the individual's previous performance. Evaluation frequently supports human resource activities such as salary administration, promotion or termination decisions, and identification and/or recognition of good or bad performance. Development and evaluation may appear interdependent (i.e., how can one develop an employee without some sort of evaluation), thus rendering separation difficult or impractical. However, previous research suggests that although developmental PA use strongly correlates with evaluative PA use, the uses are emphasized differently across organizations and differentially relate to organizational characteristics (Cleveland, Murphy, & Williams, 1989). As discussed in more detail below, development and evaluation have also been shown to differentially influence outcomes (e.g., appraisal accuracy) and rating processes (e.g., Murphy et al., 1984; Williams et al., 1985) further indicating that development and evaluation are distinct PA uses.

However, a PA can be and is often used for both development and evaluation within organizations, and some past research has examined the effects of combining these two uses. This research is discussed next.

Combining PA Uses

There are various studies specifically testing the assertions that a conflict exists when the PA is used for multiple purposes. The first, and perhaps most frequently cited research on multiple PA uses is a study conducted at General Electric Company by Meyer and colleagues (1965). Through interviews with employees and their supervisors, the researchers concluded that organizations which rely on a comprehensive annual PA to provide workers with valuable feedback and ultimately improve job performance would likely be better off with the work-planning-and-review method (WP&R: Meyer et al., 1965). The WP&R approach includes more frequent performance discussions, no summary judgments or ratings, separate salary discussions, and an increased emphasis on mutual goal planning and problem solving. This research introduced the perspective that there is conflict inherent in the traditional approach to PA, and organizations may be better served by separating PA uses (e.g., removing salary discussion from development).

However, the first empirical study of Meyer and colleagues (1965) assertion resulted in a different conclusion. Prince and Lawler (1986) primarily focused on the effects of salary discussion. Their results indicated that salary discussion during the review either had no impact or a slightly positive impact on the PA process (e.g.; participation), content (e.g.; discussion of performance attributes), and outcomes (e.g.; PA satisfaction) (Prince & Lawler, 1986). One explanation for Prince and Lawler's findings is that their manipulation was limited. The researchers only focused on the extent to which salary discussion was a part of the PA event. Therefore, although comparisons were made between whether salary discussion was the focus of the PA event or not, a more complete test of the effects of separation in PA uses would involve removing all aspects of the PA of an evaluative nature. Further, although discussing salary changes during the PA event did not negatively influence the outcome variables investigated by Prince and Lawler (1986), it is unclear whether supervisors in the study still provided the notification of salary changes (as well as other aspects of evaluation) at some other point in time. In effect, merely removing salary discussion from the annual event may not adequately separate development and evaluation. It is the separation of these constructs within the organizational context that the present research investigates.

The present research attempts to address this issue by removing all evaluative responsibility from the immediate supervisor, making evaluation the responsibility of someone else within the organization—in this case the supervisor once-removed. The idea is for the immediate supervisor to focus only on the developmental needs of his/her employees, separate from the annual appraisal. In effect, two separate processes, one for developmental purposes

the other for evaluation, are created. Arguably, by separating the PA uses in time and by person, a more complete and accurate investigation of the effects of separating PA uses is possible.

It is proposed that separating the PA uses (in time and person) would free the immediate supervisor to focus on providing employees development as needed, but separate from the actual PA event (and evaluation) and free from the role of evaluator. This is consistent with Murphy and Cleveland's (1995) social-psychological perspective where the importance of the larger appraisal context is recognized. In effect, it may be necessary to think beyond removing only one aspect of evaluation (e.g., salary discussion) from the annual event and instead separate evaluation and development completely within the organization.

To understand the possible effects of separating the developmental and evaluative PA components we draw on two research areas: cognitive processes/rating strategies and role conflict/power. Specifically, it is proposed that the incompatible processes/strategies and conflict in roles which may exist with the traditional PA should be reduced by separating the uses, thus leading to positive outcomes. This research is discussed next and specific hypotheses are offered.

Rating Processes and Strategies

It has been shown that cognitive processes and rating outcomes are different depending on the PA purpose (Bernardin & Beatty, 1984; DeNisi et al., 1984; DeNisi & Williams, 1988; Murphy et al., 1984; Williams et al., 1985). Specifically, Murphy and Cleveland (1995) suggest that if the rater has one purpose in mind and is then asked to use the PA for something else, the rater will have trouble processing the information for that alternative purpose. If raters do try to keep the multiple purposes in mind, there is a risk of failing to provide useful information for any of the purposes (Murphy & Cleveland, 1995). It has also been argued that accuracy in distinguishing between good and bad performers is independent from accuracy in distinguishing between an individual's strengths and weaknesses (Murphy, Garcia, Kerkar, Martin, & Balzer, 1982; Wiggins, 1973). As a result of these conflicts in cognitive processes and accuracy, researchers have hypothesized that the rater may attempt to balance the multiple, perhaps conflicting purposes by focusing on only one purpose and conducting the PA with that purpose in mind (Longenecker et al., 1987).

Previous research has also found that raters differentially search for performance information depending on the purpose of the appraisal (Williams et al., 1985). For example, raters seem to search for more comparative information when making evaluative decisions. A study by Zedeck & Cascio (1982) found that rater strategy varies—identical dimensions are

weighed, combined, and integrated differently—depending on the PA purpose. However, a later study was not able to support this finding (McIntyre, Smith, & Hassett, 1984). If raters do search for and/or integrate information differently depending on the PA purpose, it seems plausible that combining PA uses may be problematic. For example, seeking information for evaluative purposes may result in the rater missing useful developmental information and vice versa, ultimately leading to adverse reactions.

It has also been proposed that when designing a PA, organizations should be concerned with the type of feedback (e.g.; comparisons to other employees, areas to improve performance), frequency, quality, and who provides feedback or evaluations (Cummings & Schwab, 1973). It is possible that using a PA for multiple purposes is ineffective because the best system design may vary for each purpose (Klein, Snell, & Wexley, 1987). In other words, the optimal system for providing development may differ from the optimal system for evaluation. For example, it has been suggested that development should be provided frequently and informally, while evaluation should perhaps be provided annually or semi-annually (Meyer et al., 1965). In the present study, development was separated from the evaluative PA event with the intent of increasing developmental interactions between the employee and his/her immediate supervisor.

Previous research, therefore, suggests that there may be a conflict between PA uses in terms of rating processes and strategies. Arguably, this conflict may negatively influence employees' attitudinal reactions to the appraisal. For example, in their attempt to balance the different information collected for different PA uses raters may fail to provide any useful information (Murphy & Cleveland, 1995) or may attempt to balance the demands of multiple uses by focusing on one use in particular, perhaps disregarding the other (Longnecker et al., 1987). These actions may foster employee dissatisfaction with the PA process and outcomes. However, by separating the PA uses, the incompatibilities between the developmental and evaluative uses should be reduced. In effect, the PA process proposed here should reduce the conflict which has been shown to exist when the appraisal event is used for multiple purposes, leading to an improvement in employees' attitudinal reactions to the PA.

Hypothesis 1: Employees in the separated use PA system will report higher levels of PA satisfaction than those in the traditional system.

Role Conflict and Power

In Blau's (1963) analysis of two government agencies, he acknowledged that the more removed the supervisor, the less likely he/she apparently yielded to subordinate demands and the less lenient the evaluations. This is related to Follett's (1926) analysis regarding the conflict

faced by a supervisor when giving orders to subordinates. The immediate supervisor is often so close to the workers that he/she may not be an effective evaluator. Further, in Blau's (1963) description of the exercise of authority by supervisors, he views official power as the ultimate bureaucratic authority. This research suggests that status distinctions, while necessary for administrative functions, can weaken cooperation, feelings of association, and group cohesion. If evaluation, specifically authority over issues such as pay, promotions, and terminations, fosters defensiveness and inhibits cooperation among employees, removing this role from the supervisor may improve relations within the organization. Reducing the evaluative authority of the immediate supervisor may lead to employees' increased cooperation and ultimately feelings of association toward the supervisor.

The political aspects of PA is a perspective that coincides with the conflicting roles of the traditional evaluator. "In essence, the process of appraising does not agree with 'helping'" (Drenth, 1984: 227). We suggest that this conflict is especially evident when appraisals are used for evaluation. When assessing performance, managers are believed to avoid unnecessary conflict and minimize the extent to which administrative responsibilities create barriers between them and their subordinates; this is the result of being familiar with their workers' personal lives and needs as well as the desire to maintain a functional operating unit (Longenecker et al., 1987; Murphy & Cleveland, 1995). In an empirical investigation of the role of social and situational influences in the PA process, researchers found that supervisors' affect toward their subordinate mediated the effect of demographic similarity and supervisor-subordinate work relationship on the supervisor's PA rating of that subordinate (Judge & Ferris, 1993). If feelings toward the subordinate have a spill-over effect, or the desire to avoid conflict distorts performance ratings, the resulting PA may be less than accurate, perhaps failing to differentiate between good and bad performers. Thus, employees may not see much value in the PA as an objective evaluation. This lends further support to Hypothesis 1 above.

Criticizing employees, as is often done in PAs, is believed by many researchers to foster defensiveness and rationalization which usually result in nonconstructive responses (Blau, 1964; Meyer et al., 1965). As suggested by Drenth (1984), PA is a sensitive matter, often eliciting negative psychological responses such as resistance, denial, aggression, or discouragement particularly if the assessment is negative. These negative feelings will likely upset the relationship between the evaluator and those being evaluated (Blau, 1964; Drenth, 1984). By removing the evaluative role from the immediate supervisor, a less adversarial relationship between the employee and their supervisor may develop as status distinctions give way to cooperation. Since the evaluative and disciplinary actions become the responsibility of a

manager once-removed, the employee/supervisor relationship may become more constructive as the defensiveness and apprehension that often accompanies evaluation is no longer an issue. It is, therefore, important to assess how this change in performance management affects the employees' feelings about their supervisor.

Hypothesis 2: Employees in the separated use PA will report higher satisfaction with their supervisors than those in the traditional system.

The job of the first-line supervisor has changed drastically over the last few decades due to the growth of participative management, increased use of autonomous work groups, use of computers in management, and the influence of staff specialists (Kerr, Hill, & Broedling, 1986). Many practitioner-oriented articles suggest this very idea, asserting that leaders must now coach the development of employees' capabilities and competence, acting as facilitators or even cheerleaders (Belasco & Stayer, 1994; Lawson, 1992; McCarthy, 1991; Moskal, 1988; Shaughnessy & Howell, 1993). Kerr et al. (1986) propose that activities such as carrying out "human relations" through counseling, nurturing, and maintaining external boundaries takes much of the supervisor's time, while work planning and scheduling, coordination and control, and maintaining quality, safety, and machinery takes up less of their time. Separating the PA uses may free the immediate supervisor to focus on coaching and mentoring, as opposed to discipline and evaluation.

The above discussion suggests that the improved relationship between the supervisor and the employee will result in employees feeling at ease with their supervisor. As supervisors appear more approachable, employees should solicit their coaching and mentoring assistance more frequently and with less apprehension. Constructive discussion, planning, and utilization of development and training opportunities may result. The idea that employees in the traditional system may not wish to expose their difficulties or questions to the person who also handles their discipline and evaluation is not new (Blau, 1963). Research indicates that developmental guidance is most often sought from other resources, such as Human Resources or colleagues, when the supposed coach/mentor is in a traditional supervisory role (Blau, 1963; Fletcher and Williams, 1985; Gabor, 1990). A defective relationship between the supervisor and the employees may, therefore, prevent the employee from development and the supervisor from an intimate understanding of many work place issues. By removing the evaluative role from the immediate supervisor he/she may then become the primary provider of advice and support.

Hypothesis 3: Employees in the separated use PA will report that they are more likely to seek their supervisors' assistance with developmental issues than those in the traditional system.

As discussed above, in the traditional performance management system where raters are instructed to use the PA for multiple purposes, the ratee may focus on one purpose while the other (s) are disregarded (Gabor, 1990; Longenecker et al., 1987). Because evaluative needs are very important to the organization and often a necessity in terms of administrative decisions (i.e.; pay distribution, promotions, terminations, layoffs), development is the function likely to be disregarded. By shifting the development function to another person and time, this should no longer be the situation and employees will become more aware of job development opportunities. Moreover, practitioner oriented articles and studies have proposed that while employees are increasingly expected to take responsibility for their own career development, it is also imperative for organizations, particularly supervisors, to take an active role in assisting in and promoting development opportunities (Dennis & Axel, 1991; Gooding, 1988; Haskell, 1993). If the immediate supervisor is no longer required to act as an evaluator, they will perhaps then focus their attention on employee development. Thus,

Hypothesis 4: Employees in the separated use PA will report greater awareness of development opportunities than those in the traditional system.

On the other hand, the link between using development and receiving valued outcomes (e.g., salary increase) may become tenuous under the separated use PA. Development may no longer be viewed as instrumental to receiving outcomes because the two components (development and evaluation) are separate. Indeed, in the absence of the immediate supervisor's impact on extrinsic rewards resulting from the appraisal, employees in the treatment group may report less intent to develop in the future.

Moreover, if the immediate supervisor focuses more on development now that it is separated from evaluation, thus resulting in higher employee development, the employees in this group may not be as likely to report future use of development. Arguably, if development is overshadowed or neglected during the PA process, as proposed occurs in the traditional PA, employees may be inclined to say they do plan to use development. The control group (i.e., traditional PA) may, therefore, report more future use of development because they have not previously been provided development.

Hypothesis 5: Employees in the separated use PA will report they intend to use less development than those in the traditional system.

In sum, separating the developmental and evaluative PA uses in time as well as by person, cognitive and role conflict should be reduced. The separated use by system should influence employee attitudes regarding the appraisal, their supervisor, and development and behavioral intentions toward development.

Method

Sample

Subjects were employees of a large international manufacturing company. The facility used in this research is located in the Southern United States and manufactures and distributes production equipment. There were over 20 departments in this facility, thus the employee participants represent a mix of production/distribution workers, customer service clerks, and staff level personnel (e.g.; accountants, computer administrators). Any employees directly familiar with the research project and/or hypotheses, such as the HR Manager and Facility Vice President, were excluded from the study.

Summary statistics of the sample are reported in Table 1. Subjects were mostly female (76.5%) and white (92%). The average age was 40 years. The mean organizational and position tenure are 9.79 and 4.91 years respectively. On average, employees had been promoted 1.65 times within this company.

Prior to the start of this research project, all employees were given a formal PA once a year on their hire date. This was the time when employees were provided past performance feedback, future areas for improvement, training needs assessment, and any merit increase. The change in appraisal system was an actual organization intervention, designed in cooperation with the researchers. Top HR managers in the organization were familiarized with relevant research suggesting that separating appraisal use might favorably affect employee attitudes and behaviors. They then participated in the design of the intervention and the experimental study.

Study Design

Because the supervisor is likely in the best position to provide input concerning their employees' development (Bonito, 1990; Finn, 1991), it was determined that the supervisor once-removed would provide all evaluation. The supervisor once-removed, therefore, was to provide the annual review as well as conduct any needed disciplinary action, freeing the immediate supervisor to provide development. This was done to reduce the confusion caused by using the PA for multiple purposes (Bernardin & Beatty, 1984; DeNisi et al., 1984; DeNisi & Williams, 1988; Murphy et al., 1984; Williams et al., 1985), and to reduce the role conflict when the same person provides both development and evaluations (Blau, 1964; Latham & Wexley, 1981; Meyer et al., 1965).

Employees were assigned to one of two groups: (1) formal appraisal to be administered by the supervisor once-removed and used only for evaluation (treatment group), or (2) performance appraisal conducted as in the past, by the immediate supervisor for both

development and evaluation (control group). Both groups of employees were administered two pretest surveys, approximately two months apart, and one posttest survey after they had received their PA for that year. Of 152 employees involved in the study (again, any employee directly familiar with the study were excluded from participation), 116 completed at least one pretest (95% completed both) and the posttest survey for a response rate of 76.3%. Forty-five respondents were in the treatment group and 71 were in the control group. There was no indication of a selection-attrition bias—86% of the treatment group and 84% of the control group remained for the entire study.

It was not possible to randomly assign employees to the groups because there needed to be at least one level between the immediate supervisor and the Facility Vice President. This was necessary so that the Facility Vice President would not be burdened by having to complete a large number of PA's. Therefore, employees in the treatment group had at least two people between them and the Facility Vice President, and those in the control group generally had a supervisor who reported directly to the Facility Vice President. It should also be noted that all employees reporting to the same supervisor were assigned to the same group. Therefore, no supervisor had some employees whose PA he/she administered and other employees where the PA was administered by the supervisor once-removed. However, the different occupational groups within the organization were represented in both the treatment and the control group (e.g., some customer service clerks were assigned to the treatment group and some to the control group). Even though the two groups were fairly similar, this experiment should be termed a quasi-experiment since random assignment of groups was not feasible. Therefore, the pretest data was essential in order to detect selection-maturation and local history interactions, and to obtain an estimate of the correlation between observations taken from a single group across a known time interval (Cook & Campbell, 1979).

Each employee signed an informed consent form with a tracking number so that individual responses to the three surveys (two pretest and one posttest) could be matched. Employees were administered the first pretest survey in groups of 20-30. Only the researcher was in the room with the employees as they completed the surveys to ensure confidentiality. The surveys were kept by the researcher; no one associated with the company was ever shown individual survey responses.

At the time of the first pretest, it was not known to which group each employee would be assigned. In the two months following the first pretest, the groups were determined by the researcher and the facility's HR Manager. The employees, however, were not told of the change in the appraisal system (if they were part of the treatment group) until one week after the second

pretest. This was done so that the group membership would not affect responses to the pretests. Supervisors were informed of the PA process (depending on the group they were in) by the HR Manager. They were provided standardized details on how the PAs were to be conducted and how to explain the change to their employees (e.g., permanent change in the PA system, intent was for immediate supervisors to focus on employee development and increase standardization of appraisals).

The second pretest was administered in the same way as the first. Each employee received a survey which matched the number on their signed consent form and the first pretest. Because the researcher did not have access to the numbers corresponding to each employee and was also unfamiliar with the employees, the HR Manager distributed the surveys but left the room afterwards room to ensure confidentiality.

All immediate supervisors and supervisors once-removed were instructed by the facility's HR manager on their evaluative or developmental (or both) role depending on their group after the second pretest. The process was standardized to promote reliability of treatment implementation within groups and also to ensure that supervisors across groups provided the "correct" feedback (only evaluation, only development, or both in the case of the control group) thus increasing internal validity. Specifically, immediate supervisors in the treatment group were told that they would no longer be responsible for the annual appraisal but that since only evaluation would be provided by the supervisor once-removed during the annual PA they were to continue to provide their employees with development. Supervisors once-removed were instructed to only provide evaluative feedback (e.g., performance ratings, merit increases, promotions) during the appraisal event and that the immediate supervisor would still be responsible for employee development.

The employees in the treatment group ceased to be evaluated and disciplined by their immediate supervisor after the second pretest. All evaluations and disciplinary actions were conducted by the supervisor once-removed, and the immediate supervisor acted only in a developmental role. However, if requested by the supervisor once-removed, the immediate supervisor did provide information regarding an employee's performance. A manipulation check was, therefore, necessary to ensure that the employees in the treatment group perceived that their immediate supervisor was no longer the primary evaluator of their performance. The manipulation check will be discussed in the "Measures" section. The control group continued in the traditional system where the supervisor both evaluated and provided employee development during the annual PA.

Since PAs were scheduled on the anniversary of an employee's hire date, administration of posttest surveys varied among employees. Every three months (for one year) after the manipulation was introduced, posttest surveys were administered to those employees who had been given their PA in the prior three months. This meant that employees completed the posttest questionnaire one to three months after their actual appraisal event. Employees in both the control and treatment groups who had received a review in the previous three months were brought together in groups of 15-20 to complete the posttest.

Measures

Survey items included a combination of existing measures and items developed specifically for this study. The survey was pilot-tested with the company's HR Department to ensure that it was complete, easy to follow, and that the items were not ambiguous. A 1-7 Likert scale (1=strongly disagree, 7=strongly agree) was utilized for all survey items.

Satisfaction. Greller's Satisfaction with PA (1978) was used to measure satisfaction with the PA. Greller's scale is made up of four dimensions: utility, satisfaction, anxiety, and derogation. (Cronbach's alpha, [α]=.90). The Job Descriptive Index (Smith, Kendall, & Hulin, 1969) measured satisfaction with the supervisor (α=.82).

Development. The researchers developed the items regarding seeking supervisor's assistance with developmental issues, awareness of developmental opportunities, and the future use of development. Individual items measuring each construct were combined to create one index for each construct. There were four items concerning how likely the employee is to seek their immediate supervisor's assistance for development (e.g., "I am likely to approach my immediate supervisor regarding career development"; α=.77), two items assessing awareness of developmental opportunities (e.g., "There are very few development opportunities within this company"—reverse scored; α=.67), and two items assessing the employees' reported future use of development (e.g., "I plan to seek and use development tools in the future"; α=.74).

Manipulation Check. Questions regarding who provides the development and evaluative aspects of the PA were used to provide a manipulation check. As in a similar study on performance appraisals (Taylor et al., 1995), this is done to check that the manipulation was perceived by the employees. The questions were "my immediate supervisor is the primary evaluator of my performance," and "someone other than my supervisor handles the evaluation of my performance and discipline" (reverse-scored). The two items were averaged to create one index (α=.69).

Discussion Check. Communication between the treatment and control groups about the PA system may introduce a confounding bias in the manipulation. One question was used to

measure the degree of discussion surrounding the PA system. This allowed investigation into whether the employees discussed the PA system more after the manipulation was introduced and helped assess whether interaction between the groups confounded the results.

Control Variables. The last page of the survey contained questions regarding gender (1=male, 0=female), race (1=white, 0=non-white), job tenure (in years), and number of promotions. These items allowed for investigation of possible differences between the control and treatment groups.

Performance Rating. The organization provided the performance rating each individual received on the appraisal just prior to the research project as well as the rating since the manipulation had been implemented. These were used to investigate whether performance ratings differed between groups or as a result of the manipulation. Employees were given an overall rating in one of the following four categories: (1) below standards, (2) meets standards, (3) exceeds standards, or (4) distinguished.

Analysis

Correlations between the corresponding variables in the first and second pretests were needed to determine if observations taken from a single group across a known time interval were consistent (Cook & Campbell, 1979). The correlation approach to investigating the relationship between responses on the first and second pretests is also compatible with the test-retest estimation of reliability (Pedhazur & Pedhazur Schmelkin, 1991). A high correlation would permit the averaging of the two pretest measures to create one variable. This variable is then treated as a covariate, controlling for the initial level of the respective measure and for selection differences between the groups (Cook & Campbell, 1979).

It should be acknowledged, however, that carry-over biases from measuring the same people with the same measure twice may result in an overestimation of stability (Pedhazur & Pedhazur Schmelkin, 1991). In order to lessen any carry-over biases to the pretest measures, the time interval between the two tests was relatively large—two months. This increase in time interval may lead to low correlations between the two pretests due to the characteristics under investigation actually changing as time passes. The two month time interval is appropriate since it should not be too excessive nor insignificant; conceivably balancing carry-over problems associated with short time intervals with the probability that longer time intervals may lead to actual changes in characteristics.

In order to check for selection-maturation, mean differences between the control and treatment groups were computed for both the first and second pretests. This would determine if one group is changing on any of the measures at a rate different from the other group.

Comparing the groups' responses on the first pretest also allows detection of whether one group exhibits a higher initial level of a measure(s) compared to the other.

ANCOVAs were used to examine the effect of separating PA use on employee attitudes and behavioral intentions. First, however, ANCOVA was performed to check the effectiveness of the separated use PA manipulation. Those employees in the separated use group should report lower perceptions that evaluation is handled by their immediate supervisor. Separate ANCOVAs, using the pre-survey measure of the relevant dependent variable were then performed to test the hypotheses.

Results

Descriptive statistics for the variables used in the analysis are shown in Table 1.

TABLE 1: Descriptive Statistics and Correlations for Both Groups

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Age	39.63	9.33													
2 Number of promotions	1.65	1.99	.24*												
3 Gender (male=1)	.23	.43	-.09	-.04											
4 Race (white=1)	.92	.27	-.02	.02	-.07										
5 Job tenure	4.91	3.71	.43**	.51**	-.01	.16									
6 1st PA rating	2.43	.51	.11	.03	-.17	-.14	.11								
7 2nd PA rating	2.40	.53	.07	.08	-.04	-.15	.11	.58**							
8 PA satisfaction	3.95	1.06	.11	.17	-.12	-.10	.08	.38**	.29**						
9 Satisfaction w/ supervisor	5.00	1.02	.28**	.18	.02	-.15	.05	.19	.16	.41**					
10 Seek supervisor assistance	4.89	1.06	.16	.09	.01	-.03	.17	.17	.14	.52**	.51**				
11 Aware of development	3.32	1.16	-.02	.15	.10	-.05	.06	.08	-.02	.34**	.28**	.21*			
12 Use development	4.82	.99	-.10	.17	-.11	-.01	-.07	.01	.03	.04	.11	.11	.00		
13 Discussion check	3.35	1.51	.02	-.05	.30**	.11	-.05	-.19*	-.11	-.34**	-.07	-.18	-.13	-.01	
14 Manipulation check	5.38	1.22	.17	-.01	.15	.05	.10	.09	.11	.24*	.26**	.37**	.13	.13	.01

Note: Statistics reported here for the dependent variables are from the posttest data. ** $p < .01$; * $p < .05$

Pretest Analysis

The bivariate correlation coefficients between the first pretest responses and the corresponding second pretest responses are shown in Table 2. All dependent variables, with the exception of the manipulation check and degree of discussion about the PA system, had significant positive correlations above .60. The measures appear relatively stable.

TABLE 2: Correlation Coefficients for the First and Second Pretests

Variable	Correlation
Manipulation check	.52
Discussion check	.57
PA satisfaction	.86
Satisfaction with supervisor	.82
Seek supervisor assistance	.70
Aware of development	.73
Use of development	.64

Note: All correlations significant $p < .01$.

In addition, there were no significant differences between the treatment and control groups for any of the variables on either of the pretests (all p-values >.10). The employees' responses do not appear to be changing in one group compared to the other group prior to the introduction of the manipulation, suggesting that selection maturation may not be of concern (Cook & Campbell, 1979).

Because it may not be appropriate to scale the two separate pretest manipulation measures due to the low correlation shown in Table 2 (r=.52), both measures were maintained as distinct for use in the analysis of variance model. To be conservative, the first pretest PA discussion variable was used due to the plausibility that degree of discussion regarding the study increased as a result of the research project. For example, the comparison between initial PA discussion (pretest 1) and discussion at the time of the posttest will likely yield the higher contrast.

For all other variables, the responses to the first and second pretests were averaged to create one overall pretest measure for each variable. The corresponding variable from this pretest measure was entered as a covariate in the analysis of variance model for each dependent variable.

Demographic and Check Variables

There were no differences between the two groups on any of the demographic variables. Moreover, there were no significant differences between the groups on either the pre-study PA rating (Mc [control group]=2.44, Mt [treatment group]=2.41, F=.12, n.s.) or PA rating once the manipulation was introduced (Mc=2.41, Mt=2.39 F=.05, n.s.).

The ANCOVA results for the manipulation check show a significant difference between the control and treatment groups after controlling for the pretest manipulation. Employees in the treatment group were significantly less likely to perceive that their immediate supervisor is the person most likely to evaluate their performance (Mc=5.57, Mt=5.08, F=4.72, p<.05). In other words, employees under the new system reported significantly higher perceptions that someone other than their supervisor handles evaluation and discipline compared to the control group after controlling for the initial level of this variable. The manipulation, therefore, appears to have been effective. There were no differences between the two groups regarding the degree of discussion surrounding the PA system (Mc=3.29, Mt=3.28, F=.00, n.s.). After controlling for the initial level, it does not appear that one group is more likely to discuss the PA system compared to the other. The control and treatment group means for the manipulation, discussion check, PA rating, and outcome variables are shown in Table 3.

TABLE 3: ANCOVA Results for Control and Treatment Groups

Variable	Control Group	Treatment Group	F Value
Manipulation check	5.57	5.08	4.72*
	(.14)	(.18)	
Discussion check	3.29	3.28	.00
	(.14)	(.18)	
PA rating	2.39	2.40	.01
	(.05)	(.06)	
PA satisfaction	3.94	3.96	.01
	(.10)	(.13)	
Satisfaction with supervisor	4.98	5.03	.07
	(.11)	(.14)	
Seek supervisor assistance	4.91	4.85	.11
	(.11)	(.14)	
Aware of development	3.11	3.34	.03
	(.11)	(.14)	
Use of development	5.00	4.54	14.20**
	(.08)	(.10)	

*Note: Standard deviations in parentheses; * p < .05; ** p < .01*

Test of Hypotheses

There was no difference between the treatment and control groups regarding their satisfaction with the performance appraisal after controlling for the respondents' initial PA satisfaction level (Mc=3.94, Mt=3.94, F=.00, n.s.). Moreover, the groups were not significantly different in reported levels of satisfaction with their supervisor (Mc=4.98, Mt=5.03, F=.07, n.s). Although the treatment group reported higher supervisor satisfaction, the difference was not significant. Therefore, Hypotheses 1 and 2 were not supported.

Employees in the control group were more likely to approach their supervisor for development feedback, but the difference was not significant after controlling for the pretest measure (Mc=4.91, Mt=4.85, F=.11, n.s.). There was also no significant difference between the two groups on the variable aware of development (Mc=3.11, Mt=3.12, F=.00, n.s.). Hypotheses 3 and 4 were not supported.

Regarding future use of development, there was a significant difference between the groups (Mc=5.21, Mt=4.80, F=7.44, p<.01). Employees in the control group reported a stronger intention to use development opportunities available compared to the treatment group after controlling for the pretest measure, thus supporting Hypothesis 5.

Discussion

This study investigated the effects of a separated use PA system on employee attitudes and behavioral intention. Previous research indicates that a conflict may exist when PAs are utilized for both evaluation and development. By removing the evaluative role from the immediate supervisor and giving it to someone more removed from the situation, the supervisor once-removed, it was hypothesized that variables such as employee satisfaction with the appraisal and their supervisor, awareness of development, and employees' feelings toward their supervisors would be improved. Moreover, it was argued that under the new system employee intentions toward development may be affected.

There was a significant difference between the groups on reported future use of development. As discussed above, the separation in PA uses may have weakened the instrumentality of using development to gain valued outcomes for employees in the treatment group. It is conceivable that they were no longer motivated to use development because there was no longer a link between development and outcomes such as salary increases and promotions. On the other hand, perhaps employees in the control group were not provided development by their supervisors and, therefore, plan to get development in the future. Arguably, employees in the control group were not provided development at all because the

time when this normally occurs is during the PA and it may have been neglected due to the conflict discussed above. Therefore, the control group may have responded that they are more likely to seek and use development in the future because they have not been provided or engaged in development. This is of course speculation and should be investigated further perhaps by considering the effect on the actual use of development and not just behavioral intentions.

Although there were no significant differences between the control and treatment groups on any of the other outcome variables investigated, the results for satisfaction with supervisor and awareness of development were in the hypothesized direction. Interestingly, it appears that this substantial change in the performance management system at this organization failed to influence employee attitudes regarding the PA process, their supervisor, and development one way or the other.

Implications

Although organizations have attempted to remove development from the annual review (Devanna, Fombrun, & Tichy, 1984; McNerney, 1995), this is the first empirical investigation of such an intervention. Given previous research on PA uses and conflicts that may exist (e.g., Meyer et al., 1965; Murphy & Cleveland, 1995), investigating the effects of separating the evaluative and developmental uses within an organizational context is an important step to gaining a better understanding of the issues. Therefore, the minimal effects found in this study still contribute to our understanding of how the performance appraisal process may influence work-related variables. It is often assumed that the immediate supervisor is in the best position to appraise their employees (Kanin-Lovers, 1990; Lee, 1990). This particular study suggests that this may not necessarily be true. Direct supervisors and supervisors once-removed may be equally effective in providing reviews if employee attitudes are of interest or if these attitudes influence other organizational outcomes that are deemed important.

Further, the organization studied here indicated that they would like to decrease the number of people responsible for providing performance reviews. By having the supervisor once-removed conduct the PA rather than the immediate supervisor, the number of raters is reduced by 73% in this particular facility. This would minimize the number of people that need to be trained to conduct PAs. It is also possible that supervisors further up in the hierarchy have more experience with PAs, and thus are better raters. On the other hand, these supervisors may be less familiar with the employees' performance. They may also be unwilling to conduct additional appraisals since this would obviously require more of their time.

Limitations and Future Research

As with any field experiment, there is a difficulty in controlling intervening variables. Although one group did not necessarily discuss the PA system more than the other group, all employees may have been aware of the new system and/or believed it was an organizational experiment. A preferred study design would be to separate the two groups so as to eliminate possible awareness among the groups that employees' PAs have or have not changed. On the other hand, the results of the manipulation check suggest that this research was successful in removing the evaluative PA role from the employees' immediate supervisor. Further, the significant difference between the control and treatment groups on reported future use of development indicates that employees in the separated use PA group were in fact influenced by the manipulation.

Though employee attitudes are important both theoretically and practically, there are other variables such as appraisal accuracy, rater and supervisor attitudes, and employee and organizational performance that would be interesting to investigate. Organizations are often interested in finding ways to improve upon the accuracy of the performance rating system, as evident from the extensive research in this area (e.g.; Murphy et al., 1982; Wiggins, 1973). The separated use PA system could decrease the accuracy of the ratings since someone who is perhaps in the best position to make evaluations, the immediate supervisor, is no longer the evaluator. However, as argued above, removing evaluation from the person most familiar with the employee and their personal situation may decrease political motives underlying the inflation or deflation performance ratings. Therefore, the separated use PA system's effect on appraisal accuracy would be an interesting issue to investigate.

In addition, this research limited the analysis to employee attitudinal effects, yet other organizational members may also be affected by the system. The change in power structure incurred by removing the evaluative authority from the immediate supervisor discussed earlier is an example of the importance of investigating effects on other organizational members. It would likewise be important to investigate the effect on the supervisor once-removed. It is conceivable that they are not happy with the new system because it has increased their workload. They may also be uncomfortable rating employees that are not under their direct supervision. This may be particularly true if they are unfamiliar with the employees' performance. Although it was not feasible to assess these measures in this study due to the limited number of supervisors in each group, these would be important variables to measure in order to get a more complete picture of the effects of separating PA use.

Another deficiency in this study is related to the lack of a behavioral outcome measure. For example, human resource practices effect on employee and organizational performance is often of great interest (e.g., MacDuffie, 1995; Huselid, 1995; Welbourne & Andrews, 1996). Although employee attitudes may ultimately impact performance, it is equally important to investigate the actual performance outcomes of the new system. Moreover, other organizational outcomes such as turnover, absenteeism, occurrence of grievances, and actual use of development may be influenced and should be considered in future studies.

In conclusion, this research study has looked at separating the developmental and evaluative PA use and the effects on employee attitudes and behavioral intentions. This analysis has provided a step towards an integration of previous theory and research into a new performance management system, investigating its effects on a group of organizational members. Future research should continue in the assessment of the performance appraisal process, thus improving both our knowledge and practice.

References

Balzer, W. K., & Sulsky, L. M. 1990. Performance appraisal effectiveness and productivity. In Psychology in Organizations: Integrating Science and Practice, edited by K. Murphy & F. Saal, Hillsdale, NJ: Lawrence Erlbaum.

Belasco, J. A., & Stayer, R. C. 1994. Why empowerment doesn't empower: The bankruptcy of current paradigms. Business Horizons, 37: 29-41.

Bernardin, H. J., & Beatty, R. W. 1984. Performance Appraisal: Assessing Human Behavior at Work. Boston: Kent.

Blau, P. M. 1964. Exchange and Power in Social Life. New York: John Wiley & Sons, Inc.

Blau, P. M. 1963. The Dynamics of Bureaucracy. Chicago, IL: The University of Chicago.

Bonito, J. 1990. Motivating employees continuous improvement efforts-Part 3: Additional critical success factors. Productivity Management Review and APICS News, 10: 32-33.

Cleveland, J. N., Murphy, K. R., & Williams, R. E. 1989. Multiple uses of performance appraisal: Prevalence and correlates. Journal of Applied Psychology, 74: 130-135.

Cook, T. D., & Campbell, D. T. 1979. Quasi-Experimentation: Design & Analysis Issues for Field Settings. Chicago, IL: Rand McNally College Publishing Company.

Cummings, L. L., & Schwab, D. P. 1973. Performance in Organizations: Determinants and Appraisal. Glenview, IL: Scott, Foresman.

Deming, W. E. 1986. Out of the Crisis. Cambridge: MIT Institute for Advanced Engineering Study.

DeNisi, A. S., Cafferty, T., & Meglino, B. 1984. A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33: 360-396.

DeNisi, A. S., & Williams, K. J. 1988. Cognitive approaches to performance appraisal. In Research in Personnel and Human Resource Management (Vol. 6), edited by G. Ferris & K. Rowland, Greenwich, CT: JAI Press.

Dennis, H., & Axel, H. 1991. Encouraging Employees Self-Management in Financial and Career Planning, New York, NY: The Conference Board.

Devanna, M. A., Fombrun, C. J., & Tichy, N. M. 1984. A framework for strategic human resources management. Strategic Human Resource Management, 33-51.

Drenth, P. J. D. 1984. Personnel appraisal. In P. J. D. Drenth, Hk. Thierry, P. J. Williams, & C. J. de Wolff (Eds.), Handbook of Work and Organizational Psychology. Amsterdam: John Wiley & Sons, Ltd.

Finn, W. T. 1991. One on one coaching. Successful Meetings, 4: 102, 104.

Fletcher, C., & Williams, R. 1985. Performance Appraisal and Career Development. London: Hutchinson & Co. Ltd.

Follett, M. P. 1926. The giving of orders. In H. C. Medcalf (Ed.), Scientific Foundations of Business Administration. Baltimore: Williams & Wilkins, Co.

Gabor, C. 1990. What is so frightening about "driving out fear?" An analysis of one organization's attempts to redesign its annual performance system. In G. N. McLean, S. R. Damme, & R. A. Swanson (Eds.), Performance Appraisal: Perspectives on a Quality Management Approach. University of Minnesota Training and Development Research Center and the American Society for Training and Development Research Committee.

Gooding, G. J. 1988. Career moves—for the employee, for the organization. Personnel, 65: 112-116.

Greenberg, J. 1986. Determinants of perceived fairness of performance evaluations. Journal of Applied Psychology, 71: 340-342.

Greenberg, J. 1990. Organizational justice: Yesterday, today, and tomorrow. Journal of Management, 16: 399-432.

Greller, M. M. 1978. The nature of subordinate participation in the appraisal interview. Academy of Management Journal, 21: 646-658.

Harvey, E. L. (1995). Separating employee development from salary systems. The Human Resource Professional, 8: 3-5.

Haskell, J. R. 1993. Getting employees to take charge of their careers. Training and Development, 47: 51-54.

Huselid, M. 1995. The effects of strategic human resource management and human resource planning on firm performance. In J. Miner and D. Crane (Eds.), Advances in the Practice, Theory, and Research of Strategic Human Resource Management. New York, NY: HarperCollins.

Jawahar, I. M., & Williams, C. R. 1997. Where all the children are above average: The performance appraisal purpose effect. Personnel Psychology, 50: 905-925.

Judge, T. A. & Ferris, G. R. 1993. Social context of performance evaluation decisions. Academy of Management Journal, 36: 80-105.

Kanin-Lovers, J. 1990. Making performance evaluation work. Journal of Compensation and Benefits, 5: 360-362.

Kerr, S., Hill, K. D., & Broedling, L. 1986. The first-line supervisor: Phasing out or here to stay. Academy of Management Review, 11: 103-117.

Klein, H. J., Snell, S. A., & Wexley, K. N. 1987. Systems model of the performance appraisal interview process. Industrial Relations, 26: 267-280.

Latham, G. P. 1988. Human resources training and development. Annual Review of Psychology, 39: 545-582.

Lawson, K. E. 1992. First you train, then you coach. Bottomline, 9: 34-35.

Lee, C. (1990). Smoothing out appraisal systems. HRMagazine, 35: 72-70.

Lind, E. A., & Tyler, T. R. 1988. The Social Psychology of Procedural Justice. New York: Plenum.

Longenecker, C. O., Gioia, D. A., & Sims, H. P., Jr. 1987. Behind the mask: The politics of employee appraisal. The Academy of Management Executive, 1: 183-193.

MacDuffie, J. P. 1995. Human resource bundles and manufacturing performance: Organizational logic and flexible production systems in the world auto industry. Industrial and Labor Relations Review, 48: 197-221.

McIntyre, R. M., Smith, D. E., & Hassett, C. E. 1984. Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69: 147-156.

McNerney, D. J. 1995. Improved performance appraisals: Process of elimination. HRFocus, 72: 1, 4-5.

Meyer, H. H., Kay, E., & French, J. R. P., Jr. 1965. Split roles in performance appraisal. Harvard Business Review, 43: 123-129.

Moskal, B. S. 1988. Managing in the '90's: Tomorrow's best managers: Where are they now? Industry Week, 237: 32-34.

Murphy, K. R., Balzer, W. K., Kellam, K. L., & Armstrong, J. 1984. Effect of purpose of rating on accuracy in observing teacher behavior and evaluating teacher performance. Journal of Educational Psychology, 76: 45-54.

Murphy, K. R. & Cleveland, J. N. 1995. Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives. Thousand Oaks, CA: Sage Publications.

Murphy, K. R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W. K. 1982. Relationship between observational accuracy and accuracy in evaluating performance. Journal of Applied Psychology, 67: 320-325.

Ostroff, C. 1993. Rater perceptions, satisfaction and performance rating. Journal of Occupational and Organizational Psychology, 66: 345-356.

Pedhazur, E. J., & Pedhazur Schmelkin, L. 1991. Measurement, Design, and Analysis: An Integrated Approach. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Prince, J. B., & Lawler, E. E. 1986. Does salary discussion hurt the developmental performance appraisal? Organizational Behavior and Human Resource Decision Processes, 37: 357-375.

Rudin, J. P., & Boudreau, J. W. 1996. Information acquisition in promotion decisions. Human Relations, 49: 313-325.

Shaughnessy, C., & Howell, M. 1993. Coaching and the long-distance manager. Canadian Banker, 100: 40-43.

Smith, P. C., Kendall, L. M., & Hulin, C. L. 1969. The Measurement of Satisfaction in Work and Retirement: A Strategy for the Study of Attitudes. Skokie, IL: Rand-McNally.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. 1993. Applicant reactions to selection procedures. Personnel Psychology, 46: 49-76.

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.

Vroom, V. H. (1964). Work and Motivation. New York: Wiley.

Welbourne, T. M., & Andrews, A. O. 1996. Predicting the performance of initial public offerings: Should human resource management be in the equation? Academy of Management Journal, 39: 891-919.

Wiggins, J. S. 1973. Personality and Prediction: Principles of Personality Measurement. Reading, MA: Addison-Wesley.

Williams, K. J., DeNisi, A. S., Blencoe, A. G., & Cafferty, T. P. 1985. The role of appraisal purposes: Effects of purpose on information acquisition and utilization. Organizational Behavior and Human Resource Decision Processes, 35: 314-339.

Zedeck, S., & Cascio, W. F. 1982. Performance appraisal decisions as a function of rater training and purpose of the appraisal. Journal of Applied Psychology, 67: 752-758.