The Current State of Performance Appraisal Research and Practice: Concerns, Directions, and Implications

Robert D. Bretz, Jr.
George T. Milkovich
Cornell University
Walter Read
Empire Blue Cross and Blue Shield

On the surface, it is not readily apparent how some performance appraisal research issues inform performance appraisal practice. Because performance appraisal is an applied topic, it is useful to periodically consider the current state of performance research and its relation to performance appraisal practice. This review examines the performance appraisal literature published in both academic and practitioner outlets between 1985 and 1990, briefly discusses the current state of performance appraisal practice, highlights the juxtaposition of research and practice, and suggests directions for further research.

According to some writers, performance appraisal research has done little to improve its usefulness as a managerial decision-making tool (Banks & Murphy, 1985; Napier & Latham, 1986; Thorndike, 1949). Some have suggested that the issues dominating performance appraisal research (i.e. formats, evaluator training, and cognitive processing), and the methodological designs being used in this research, seem at odds with organizational realities. For example, Banks and Murphy (1985) warned that if cognitive process research continued along contemporary lines, the apparent gap between performance appraisal research and practice would increase. Napier and Latham (1986) suggested that progress on performance appraisal practice has lagged because the research that might inform practice has ignored Thorndike’s (1949) call for practicality in its quest for measurement elegance. Bernardin and Villanova (1986) concluded that better understanding of the organizational contexts in which appraisal takes place was neces-

The authors gratefully acknowledge support from the Center for Advanced Human Resource Studies, School of Industrial and Labor Relations, Cornell University, and from IBM Corporation. We thank Caroline Weber for her assistance in collecting data about the state of current performance appraisal practice. Finally, we thank the anonymous reviewers for their constructive comments on an earlier draft of this manuscript.

Address all correspondence to Robert D. Bretz, 393 Ives Hall, Center for Advanced Human Resource Studies, New York State School of Industrial and Labor Relations, Cornell University, Ithaca, NY 14853-3901.

Copyright 1992 by the Southern Management Association 0149-2063/92/$2.00.
sary in order to improve the degree to which performance appraisal research contributes to performance appraisal practice.

There is a growing concern that much organizational research, though methodologically sophisticated, lacks substantive application and is directed toward increasingly selective audiences of researchers, to the neglect of other audiences such as policy makers and managers (Bedian, 1989). This statement is representative of some, though certainly not all, of the recent performance appraisal literature. We believe that performance appraisal research can be evaluated both in terms of its theoretical contribution and its ability to inform practice. Because the rating process involves complex cognitive processes, basic research that defines the nature of the phenomenon is clearly needed. However, because performance appraisals occur in applied social and political contexts, it also is wise to consider the degree to which research is informing practice. Doing so should provide the opportunity for both researchers and managers to assess their understanding of the phenomenon. This review is undertaken in that spirit. It (a) summarizes the performance appraisal research published from 1985 through 1990, (b) provides an overview of the current state of performance appraisal practice in U.S. organizations, (c) compares the trends in performance appraisal research with the issues emerging in practice, and (d) proposes research to address what appear to be the under-studied or overlooked issues.

Recent Performance Appraisal Research

Published articles about the performance appraisal process were identified by means of a computerized literature search (ABI/Info) augmented by a review of the tables of contents of several academic and practitioner journals. This literature review is not exhaustive because it does not include technical reports, dissertations, textbooks, or chapters. However, we believe that it serves to indicate, with some precision, the focus of performance appraisal research and the manner in which appraisal researchers have chosen to allocate their limited resources. Readers interested in reviews covering earlier time periods are referred to Bernardin and Beatty, 1984; Bernardin and Villanova, 1986; DeNisi, Cafferty, and Meglino, 1984; DeNisi and Williams, 1988; Feldman, 1981; Landy and Farr, 1980; and Wexley and Klimoski, 1984. Readers should also note that some of the cognitive processing studies identified below are discussed in some detail in Lord and Maher’s (1989) review of the cognitive processing literature.

Overview of Recent Literature

The appendix serves as a guide to performance appraisal research published during this period. The appendix provides an indication of the issues being studied and the methodologies being employed. It should also assist those interested in locating research about particular topics. Recent research has been heavily weighted toward cognitive process issues. The vast majority of these studies were conducted in laboratory settings using student subjects and either paper people or video-tape formats. Rater/ratee characteristics also received considerable attention, but the research was not concentrated on any particular characteristic. Study of psychometric issues remained common, with more attention focused on halo than on other issues. Feedback issues were the rule rather than the exception. Other issues, including, formats, fairness, and appraisal procedures, however, these issues were frequently the object of research. Most often in case-study descriptions, these major segments of the literature preclude the discussion of even the key issues examined and the methods used.

Cognitive Processing of Information

Information processing issues are a major research area. Laboratory settings and studies with actual subjects and field settings in research by Banks and Murphy, 1987; Huber, Podsakoff, & Todorow, 1987; Mount & Thompson, 1987; Schmitt & Murray, 1987. Processing research concentrated around the cognitive knowledge of prior performance levels and (b) the role of memory in the process.

Raters’ knowledge of prior performance by framing or anchoring cues (Thomson & Banks, 1987). Laboratory research indicated confirmatory contrast effects (i.e., bias away from similar effects (Murphy, Balzer, & Buda, 1988). Additionally, Steinmetz (1987) found that good and poor performance levels should be presented to the raters and that raters biased their judgments toward the good level (usually good or poor) toward the good.

Raters’ expectations also may influence performance ratings. Mount and Thompson (1987) found a strong relationship between subordinates’ ratings of managers and subordinates’ expectations were congruent with prior expectations. Longitudinal study of 49 banking supervisors’ expectations introduced a confirmatory bias of prior expectations about ratee performance apparently a preexisting condition that raters did not attempt to test to the degree to which raters agreed with the expectation and whether raters agreed with or disagreed with the expectation.
Performance appraisal research encompasses organizational research, though methodologies and is directed toward interests, to the neglect of other audiences (Ali, 1989). This statement is representative of recent performance appraisal literature. Research can be evaluated both in ability to inform practice. Because the processes, basic research that defines needed. However, because performance contexts, it also is wise to consider practice. Doing so should provide the people to assess their understanding of in that spirit. It (a) summarizes the from 1985 through 1990, (b) provides appraisal practice in U.S. organizational appraisal research with the issues arch to address what appear to be the main issues.

Performance Appraisal

Appraisal process were identified by ABI/Inform augmented by a review and practitioner journals. This literature does not include technical reports, dissertations, and does not believe that it serves to indicate, with appraisal research and the manner in which their limited resources. Readers' periods are referred to Bernardin (1986; DeNisi, Cappelli, and Meighan, 1981; Landy and Farr, 1980; and also note that some of the cognitive issues in some detail in Lord and Mednick literature.

Performance appraisal research published in the Journal of Applied Psychology, 1991) also noted the need for more attention focused on halo than on other issues. Feedback issues were the one area where field studies were the rule rather than the exception. Research about sources of ratings, rater training, formats, fairness, and appraisal uses and consequences was limited. However, these issues were frequently discussed in the practitioner-oriented outlets, most often in case-study descriptions or in "how to" articles. In the following sections, these major segments of the literature are examined. Although space limitations preclude the discussion of every study, each section attempts to highlight the key issues examined and the methodologies commonly used.

Cognitive Processing of Information

Information processing issues dominated recent performance appraisal research. Laboratory settings and student subjects were very common; only a few studies used Banks and Murphy's (1985) call for incorporating non-student subjects and field settings in research designs of cognitive process issues (Hogan, 1987; Huber, Podsakoff, & Todor, 1986; Jolly, Reynolds, & Slocum, 1988; Mount & Thompson, 1987; Schmitt, Noe, & Gottschalk, 1986). Cognitive processing research concentrated around two issues: (a) how prior expectations or knowledge of prior performance levels affect the way information is processed, and (b) the role of memory in the rating process.

Raters' knowledge of prior performance appears to affect information processing by framing or anchoring current judgments (Huber, Neale, & Northcraft, 1987). Laboratory research indicated that knowledge of prior performance caused contrast effects (i.e., bias away from level of prior performance) rather than assimilation effects (Murphy, Balzer, Lockhart & Eisenman, 1985; Smith, Reilly, & Buda, 1988). Additionally, Steiner and Rain (1989) reported that the order in which good and poor performance was observed affected performance ratings and that raters biased their judgments about inconsistent extreme performance (unusually good or poor) toward the general impression already held.

Raters' expectations also may introduce bias into the rating process. For example, Mount and Thompson (1987) examined the effect of prior expectations on subordinates' ratings of managers whose behaviors were either congruent or incongruent with prior expectations. Results indicated that when behavior was congruent with expectations, appraisal results were more accurate. Similarly, in a longitudinal study of 49 banking supervisor-subordinate dyads, Hogan (1987) found supervisors' expectations introduced error into the rating process, and that disconfirmation of prior expectations appeared to lower ratings. However, consistency of rater performance apparently affects rater ability to form general impressions and categorize information. Padgett and Ilgen (1989) demonstrated that consistent rater performance led to greater use of categorization but inconsistent performance led to greater retention of behavioral information.

Moreover, it appears that job and rater knowledge also affect how information is processed. Schmitt, Noe, and Gottschalk (1986) studied 153 school administrators to test the degree to which raters used similar methods of combining information, and whether rater agreement was based on job-relevant inputs or on shared bias. They reported that overall ratings from different sources varied because different rater groups attached higher relative weights to the job-related performance.
dimensions that were most salient to them. Laboratory research using student subjects and paper people also suggested that job and rate knowledge had significant effects on conceptual similarity and rating covariance as well as on halo (Kozlowski, Kirsch, & Chao, 1986).

The role of memory has also been important in recent cognitive processing research. Virtually all of this research was conducted in laboratory settings with student subjects. Memory decay introduces bias into the rating process. For example, Kozlowski and Kirsch (1987) suggested that memory decay affected the ability to recall job and rate information and resulted in halo error and subsequently inaccurate ratings. Under laboratory conditions, ratings recorded one day after performance was observed were already affected by memory decay (Murphy & Balzer, 1986). When the rater’s memory demands are great, bias in favor of general impressions or recent performance may be expected (Murphy, Gannett, Herr, & Chen, 1986). Stress has been shown to impact memory by (a) causing less differentiation across dimensions (halo), (b) affecting information retrieval, and (c) possibly affecting categorization as well (Srinivas & Motowidlo, 1987). However, rater characteristics may moderate the degree to which memory decay is problematic. For example, Smither and Reilly (1987) concluded that rater intelligence, not rating delays, affected rating accuracy.

Other cognitive processing studies are very difficult to classify, but illustrate the diversity of this research. For example, information collected from 22 nursing supervisors was used to construct a cognitive map of their appraisal processes (Jolly, Reynolds, & Slocum, 1988). Results suggested that values accounted for significant variation in performance ratings. In laboratory settings using student subjects, personality theory (traits) influenced even behavior-based ratings (Krzystofiak, Cardy, & Newman, 1988), and information acquisition patterns (ranking versus rating) affected how the information was processed (Williams, DeNisi, Megleno, & Cafferty, 1986). Sex-role stereotypes did not affect causal attributions of performance and therefore had only small effects on performance ratings (Kinicki & Griffeth, 1985). Williams, DeNisi, Blencoe, and Cafferty (1985) reported that appraisal purpose and outcome had limited effects on how raters used information, and Feldman, Camburn, and Gatti (1986) suggested that illusory correlation was not likely to cause bias in ratings. Nathan & Alexander (1985) suggested a model for inferential accuracy based on the degree of congruence between the rater’s implicit theory of performance and the actual occurrence of behavior, and the rater’s willingness to make judgments with limited information. Finally, Dipboye (1985) reported that overemphasis on cognitive determinants of performance ratings has led to neglect of behavioral, social, and affective determinants of bias in the rating process.

Rater/Ratee Personal Characteristics

Research on sex/gender effects has yielded conflicting results. For example, no sex (or race) effects were reported in field settings where job analysis was used to develop a task-based performance appraisal instrument (Thompson & Thompson, 1985), and no gender differences were reported when rating familiar tasks in work situations where feedback was available (Shore & Thornton, 1986). Conversely, students tended to give women (Cardy, & Truxillo, 1988), and raters had tended to be less accurate when ratings (developmental) purposes. In an experiment demonstrated that females were more performance appraisals and feedback self-interestedly account for an “explanations.

Ratee age received limited research conclusions that younger subordinates performing the same job, and that supervisors of subordinate’s age (Ferris, Yair, & Field setting, Lawrence (1988) found that ratings with performance ratings. Management received higher ratings whereas those bings. However, meta analysis results were unrelated (McEvoy & Cascio, 1989).

A meta analysis of rate effects on higher ratings (Kraiger & Ford, 1985). (a) emotional disability was shown to influences of non-managerial women, defining no effect on men (Heilman & Stone) acquaintance may depend on rating. Rater affect also appeared to influence. However, Cardy & Dobbins (1986) attributed by increasing leniency but by introducing imprecision.

Rating Errors and Accuracy

The effect of rating errors on appraisals resources. Much of this recent research examined methods of measuring halo (Pulakos, 1988; Feldman, 1986). Murphy and Boudreau (1988) correlated with greater accuracy and speculate that schemas that correctly classify the relative noise. Nathan and Tippins (1990) between halo and accuracy, but Fisicaro whoship exists. However, Becker and Carpenter between halo and accuracy was ambiguous. Halo may yield either similar or divergent effects of rating (e.g. Wherry, 1983), halo may increase the performance of professors incrementally, but cannot be negated. Based on meta-analytic results, M.
em. Laboratory research using student that job and ratee knowledge had signif-
icant in recent cognitive processing re-
duction in laboratory settings with sub-
jects into the rating process. For exam-
ple, memory decay affected the and resulted in halo error and subse-
quent conditions, ratings recorded one day may have been memory decay (Mur-
phy & Gattis, 1986). The impact of the memory demands are great, bias in favor of memory decay is likely (1987) concluded that rater intelligence is difficult to classify, but illustrate the amount of information collected from 22 nursing staff and obtained from their appraisal processes as suggested that values accounted for various. In laboratory settings using student ratings, even behaviorally-based ratings of information acquisition patterns in performance. Williams, et al. (1987) did not affect causal attributions, only small effects on performance. DeNisi, Blencoe, and Cafferty d outcome had limited effects on how, and Gatti (1986) suggested that bias in ratings. Nathan & Alexander ncuracy based on the degree of congruence and the actual occurrence of judgments with limited information, overemphasis on cognitive determinants of behavioral, social, and affective

versely, students tended to give women professors higher ratings (Dobbins, Cardy, & Truxillo, 1988), and raters holding traditional stereotypes of women tended to be less accurate when ratings were made for administrative (versus developmental) purposes. In an experimental setting, Benedict and Levine (1988) demonstrated that females were more lenient with poor performers and delayed performance appraisals and feedback sessions. However, using data from within-subject analyses Pulakos, White, Oppler, and Borman (1989) concluded that gender and race account for an “extremely small” amount of variance in ratings.

Ratee age received limited research attention. A field study of nursing supervisors reported that younger subordinates were rated higher than older subordinates performing the same job, and that supervisors’ causal attributions appeared to be related to subordinates’ age (Ferris, Yates, Gilmore, & Rowland, 1985). Also in a field setting, Lawrence (1988) found that deviation from age norms was associated with performance ratings. Managers who were ahead of their age cohort received higher ratings whereas those behind their age cohort received lower ratings. However, meta-analysis results suggest that job performance and age are unrelated (McEvoy & Cascio, 1989).

A meta-analysis of race effects confirmed that rater/ratee similarity results in higher ratings (Kraiger & Ford, 1985). In studies of other personal characteristics, (a) emotional disability was shown to inflate ratings when clear professional standards were not present (Czajka & DeNisi, 1988); (b) attractiveness inflated ratings of non-managerial women, deflated ratings of managerial women, and had no effect on men (Heilman & Stopeck, 1985); and (c) the effects of rater-ratee acquaintance may depend on rating format (Kingsstrom & Mainstone, 1985).

Rater affect also appeared to influence rating behavior (Tsui & Barry, 1986). However, Cardy & Dobbins (1986) suggested that affect influenced the rating not by increasing leniency but by introducing noise into the process.

Rating Errors and Accuracy

The effect of rating errors on appraisal accuracy continues to attract research resources. Much of this recent research has examined and contrasted competing methods of measuring halo (Pulakos, Schmitt, & Ostroff, 1986; Lance & Woehr, 1986; Feldman, 1986). Murphy and Balzer (1986) reported that halo was associated with greater accuracy and speculated that this may be due to categorization schemas that correctly classify the relevant behavioral information and eliminate the noise. Nathan and Tippins (1990) also reported a positive relationship between halo and accuracy, but Fischler (1988) concluded that a negative relationship exists. However, Becker and Cardy (1986) argued that the relationship between halo and accuracy was ambiguous, that variance and correlational forms of halo may yield either similar or divergent results. Seeming to contradict theories of rating (e.g. Wherry, 1983), halo increased as the opportunity for students to observe performance of professors increased (Jacobs & Kozlowski, 1985), and true halo (true correlations between performance dimensions) was shown to have only a small effect on observed halo (Murphy & Reynolds, 1988).

Based on meta-analytic results, Murphy and Balzer (1989) concluded that the
correlation between rating errors and accuracy was very near zero and, therefore, error measures were not good indicators of rating accuracy. Because most performance is multidimensional, some correlation between performance dimensions is expected. Therefore, raters with large observed correlations may, in fact, be accurately rating performance rather than committing halo error. The ambiguity of the relationship may be due to different conceptualizations of accuracy. Lord (1985) proposed that accuracy might be better understood by distinguishing between classification and behavioral accuracy. Classification accuracy refers to the rater’s ability to place rates into cognitive categories and recall attributes that are most representative of the category prototype. Behavioral accuracy refers to the rater’s ability to recall actual ratee behavior and recognize prototypical characteristics that do not necessarily describe the ratee. It seems, therefore, that rating errors such as halo might be positively related to classification accuracy but negatively related to behavioral accuracy.

In other articles on psychometric issues, Smither, Barry, and Reilly (1989) investigated the validity of expert true score estimates, and reported that experts were more accurate than non-experts regardless of the true intercorrelations between performance dimensions. Participation in selection also affected ratings. Specifically, Schoorman (1988) reported that supervisors who had a say in the hiring decision and who viewed the applicant as favorable, subsequently tended to give lenient performance ratings, whereas those who participated in hiring but viewed the applicant as unfavorable, tended to give more severe ratings. Finally, Sulsky and Balzer (1988) argued that accuracy in performance measurement was lacking due to poor definitions of accuracy, methodological and theoretical limitations of true score development, and the absence of a cohesive theory of performance.

**Appraisal Sources**

The usefulness of self-appraisals may be affected by rating purpose, but conflicting results have been reported. Both laboratory and field studies have concluded that when used for evaluative purposes, self-appraisals were susceptible to leniency bias, but leniency decreased when appraisals were expected to be validated (Farh & Werbel, 1986; Farh, Werbel & Bedeian, 1988). Conversely, Fox and Dimur (1988) reported low validity of self-ratings regardless of the expectation of validation. Campbell and Lee (1988) suggested that self-appraisals were best suited for developmental rather than evaluative purposes and that self-appraisals can improve future performance by creating a self-fulfilling prophecy. Vance, MacCallum, Coover, and Hedge (1988) reported that among a sample of jet engine mechanics, peer, self, and supervisory ratings were equally valid sources but Fox, Ben-Nahum, and Yimon (1989) concluded that rating accuracy was positively related to rater-ratee similarity. Meta-analytic results suggested only moderate relationships exist between self-supervisor and self-peer ratings (Harris & Schaumbroek, 1988).

**Appraisal Feedback**

Most of the articles addressing feedback were conducted in field settings, distinguishing this area of research from the student subjects. Many of these studies used feedback. Discussion of pay and advancement session was shown to lead to higher entry did not influence future performance (Helmreich, Prince and Lawler (1986), report that the appraisal interview had either no relationship to behavior. However, Pearce and Porter (1980) an employee as “satisfactory” (as compared to reduced organizational commitment to the appraisal system.

Using a field study to examine feedback, it was reported that self-generated and specific and general feedback were positively related. Bannister’s (1986) experimental results indicated that message content influenced recipient response, and Klimkoski (1989) reported that feedback on performance but feedback from self and others explored feedback content in a laboratory setting. Quantity led to higher quantity, feedback quantities feedback about both led to both. Message cognition. Specifically, raters do not like scripts to be too short (Dugan, 1989).

The dimensionality of feedback also effects study of university employees Dorfman (1987) noted of performance appraisal feedback (being discussed pay and advancement) and reported that satisfaction with feedback for those who are not satisfied with feedback it perhaps may be a function of satisfaction with the content of feedback.

**Rater Training**

Recent research on rater training indicated that training focused on reducing leniency and halo but also reduced simply error reduction. In a laboratory setting, Athey and McIntyre (1986) reported that training improved retention of information, and recent literature review of 24 rater training best suited for reducing Halo, reducing errors (1986). Banks and Roberson (1985) argued that corporate standard dimensions of test of appraisals as “tests” violates accepted


Performance Appraisal

Performance appraisal was very near zero and, therefore, if rating accuracy. Because most perfor-

tion between performance dimensions is r derived correlations may, in fact, be accum-

ulating halo error. The ambiguity of the e
tualizations of accuracy. Lord (1985) u

derstood by distinguishing between s

sification accuracy refers to the rater’s
doies and recall attributes that are most b

behavioral accuracy refers to the rater’s r

recognize prototypical characteristics. It s

ems, therefore, that rating errors in c

lassification accuracy but negatively s

in Smither, Barry, and Reilly (1989) in-

one estimates, and reported that experts e

less of the true intercorrelations be-

tion in selection also affected ratings, t

supervisors who had a say in the at as favorable, subsequently tended to as those who participated in hiring but did to give more severe ratings. Finally, c

acy in performance measurement was s

ethodological and theoretical limi-

sence of a cohesive theory of perfor-

e affected by rating purpose, but con-

laboratory and field studies have con-

ces, self-appraisals were susceptible to appraisals were expected to be valid i

& Bedeian, 1988). Conversely, Fox s

self-ratings regardless of the expecta-

 suggested that self-appraisals were e

valuate purposes and that self-

by creating a self-fulfilling prophecy. e

1988) reported that among a sample of 

soratory ratings were equally valid 1989) concluded that rating accuracy. Meta-analytic results suggested s

self-supervisor and self-peer ratings were conducted in field settings, dis-

uing this area of research from those dominated by laboratory settings and student subjects. Many of these studies focused on the effects of performance feedback. Discussion of pay and advancement during the performance feedback session was shown to lead to higher employee satisfaction with the process but did not influence future performance (Dorfman, Stephan & Loveland, 1986). In contrast, Prince and Lawler (1986), reported that salary discussions during the appraisal interview had either no relationship or a positive relationship with future behavior. However, Pearce and Porter (1986), reported that feedback describing an employee as “satisfactory” (as compared to above average or outstanding) led to reduced organizational commitment and negative attitudes toward the performance appraisal system.

Using a field study to examine feedback source and message, Earley (1988) reported that self-generated and specific feedback (versus supervisor-generated and general feedback) were positively related to performance. This agrees with Bannister’s (1986) experimental results concluding that source credibility and message content influenced recipient response to the feedback. In fact, Becker and Klimoski (1989) reported that feedback from supervisors led to increased performance but feedback from self and peers did not. Ilgen and Moore (1987) explored feedback content in a laboratory setting and found that feedback about quantity led to higher quantity, feedback about quality led to higher quality and feedback about both led to both. Message content apparently also affects rater cognition. Specifically, raters do not like to give negative feedback (Larson, 1989) and are likely to rely on scripts to deliver feedback about poor performance (Dugan, 1989).

The dimensionality of feedback also has been examined. In a longitudinal study of university employees Dorfman et al. (1986) identified three dimensions of performance appraisal feedback (being supportive, emphasizing improvement, and discussing pay and advancement). Furthermore, Russell and Goode (1988) reported that satisfaction with feedback also may be multi-dimensional. Therefore, individuals who are satisfied with the performance appraisal in general, may not be satisfied with the feedback it provides. Rather, satisfaction with feedback may be a function of satisfaction with the supervisor and/or the rating received.

Rater Training

Recent research on rater training has been limited. Hedge and Kavanagh (1988) reported that training focused on minimizing rating errors successfully reduced leniency and halo but also reduced accuracy. They concluded that rater training should emphasize observation and decision-making processes rather than simply error reduction. In a laboratory study using student subjects and videotaped lectures, Athey and McIntyre (1987) showed that frame-of-reference training improved retention of information, improved accuracy, and decreased halo. A recent literature review of 24 rater training studies suggested the training methods best suited for reducing halo, reducing leniency, and improving accuracy (Smith, 1986). Banks and Roberson (1985) argued that rater training programs did not incorporate standard dimensions of test development and the notion of performance appraisals as “tests” violates accepted standards of test construction.
Performance Appraisal Formats

Of the studies examining formats, behaviorally based methods received the most, but limited attention. Using a sample of mechanics, Hughes and Prien (1986) evaluated alternative scoring methods for mixed standard scales. They found few differences between the methods and suggested choosing one based on ease of application or explanation. Prien and Hughes (1987), using a state government sample, showed that mixed standard scales can be used to identify and minimize individual rater error and system-wide problems. In 1987, Murphy and Constan concluded that behavioral anchors may lead to biased recall of performance. But 2 years later, Murphy suggested that the earlier results were not likely to be observed in organizational settings (Murphy & Pardafy, 1989).

Other Research Issues

Other research-based articles were the sole examples of studies on particular issues. For example, Barrett and Kennan (1987) reviewed court cases since Brito vs. Zia arising from terminations based on performance appraisal, and Miller, Kaspin, and Schuster (1990) discussed performance appraisal practices related to age discrimination cases. Using a managerial sample, Greenberg (1986) reported that perceived fairness of performance evaluations depended on the presence of procedural characteristics (e.g., communication, appeals process, job knowledge, consistency) and distributive characteristics (e.g., rating based on performance, action based on rating). Sackett, Zedack, and Fogli (1998) used a sample of supermarket cashiers to explore differences between typical and maximum job performance. They found a low correlation between the two and commented on the appropriateness of using procedures that tap maximum versus typical performance. Wayne and Ferris (1990) reported that political influence behavior can affect supervisor affectivity and result in more favorable performance ratings than would otherwise be expected, but the type of influence tactic used may affect the performance outcome (Kipnis & Schmidt, 1988). Finally, Napier and Latham (1986) reported that managers (a) perceived no consequences (good or bad) from conducting thorough performance appraisals, and (b) saw little practical value in doing so.

Case Studies and “How-To” Articles

The professional journals are replete with articles discussing the performance appraisal practices in various organizations and under varying conditions. Case studies and “how to” articles are common. For example, Gellerman and Hodgson (1988) discussed how American Cyanamid Company transformed a 10-level forced distribution performance appraisal system into a 3-level system that was deemed to be more consistent with the organization’s culture, and Scherkenbach (1985) explained how Ford Motor Company revised their appraisal practice to fit its focus on total quality. Other case studies include reports of the appraisal practices at Xerox (Deets & Tyler, 1986), Control Data (Gomez-Mejia, Page, & Tornow, 1985), Merck (Wagel, 1987), many federal and state agencies (e.g. Glen, 1990; Goodell, 1988; Hall, 1987; Laumeyer & Beebe, 1988), and several uniden-
haviorally based methods received the nple of mechanics, Hughes and Prien methods for mixed standard scales. They is and suggested choosing one based on Hughes (1987), using a state governs scales can be used to identify and min-wide problems. In 1987, Murphy and ors may lead to biased recall of perfor-med that the earlier results were not likely Murphy & Pardaffy, 1989).


The practitioner literature also contains several “recipes” for insuring the effectiveness of various appraisal practices. For example, advice is offered regarding how to construct and implement effective appraisal systems (Levy, 1989; Schneier, Beatty, & Baird, 1986a, 1986b), how to make effective use of team appraisals (Edwards & Sproull, 1985; Lanza, 1985), and how to make appraisals more objective (Friedman, 1986; Regel & Hollmann, 1987). One major focus in the practitioner literature addresses transforming performance appraisal from an event to a process. Advice for how to do so is typically discussed under the rubric of “performance management” (Day, 1989; Kirkpatrick, 1986; Levy, 1989; Romanoff, 1989; Schneier, 1989; Schneier, Beatty, & Baird, 1986a, 1986b).

Literature Summary

Cognitive processing issues clearly dominated this period of performance appraisal research. Prior expectations, prior job knowledge, and memory decay were all found to affect performance appraisals. The characteristics of raters and ratees, particularly the effects of gender, also received research attention. However no consensus emerged. For example, no gender (or race) effects were reported in fields studies, but student subjects in laboratory settings did exhibit gender bias. Halo and accuracy of appraisals were the psychometric topics of choice among researchers. As with gender effects, the relationship between halo and accuracy seems unresolved. Methodologically, assessment of paper people or video scenarios with student subjects in laboratory settings is the norm. The clear exceptions were the field studies of the consequences and dimensionality of appraisal feedback. It appears that salary discussions during feedback have either no effect or a positive effect on future performance, but labelling someone as satisfactory rather than above average or outstanding reduces commitment and satisfaction with the appraisal system. Some very interesting recent research was found in the “sole example” studies. These include Greenberg’s (1986) study of perceived fairness of appraisals as a function of procedural and distributive characteristics, Sackett, Zedeck and Fogli’s (1988) use of typical versus maximum job performance and Napier & Latham’s (1986) finding that managers perceived no consequences or practical value from conducting appraisals. These studies point to important issues that have barely been addressed.

The conclusion we draw from this and earlier reviews of appraisal research is that our knowledge of the rating process has expanded greatly in recent years but remains fragmented. This fragmentation appears to be caused by fundamental differences between the measurement aspects of appraisal research and the organizational purposes of performance appraisal. From a measurement perspective, the necessity to isolate specific effects has resulted in single-issuе studies conducted in laboratory settings. Moreover, most of this research addresses the consistency, not necessarily the relevance, of the measurement. The effects of prior expectations, prior knowledge of performance, and memory decay have been studied separately from the alternative uses of appraisals (administrative or development-
tal), the characteristics of raters/ratees, or the types of scales and formats employed.

Furthermore, certain appraisal issues have received considerable attention but others have been virtually ignored. 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 or the cognitive processes of the rater, how raters actually appraise performance, how they use appraisal information, or what issues they believe are important. Moreover, research is only beginning to address how context affects employee perceptions of appraisal, their reactions to appraisal outcomes, and how appraisal purpose (administrative versus developmental) moderates these relationships. Therefore, in order to understand the organizational contexts in which appraisals are made, we turn to a brief examination of the current state of performance appraisal practice.

Current Performance Appraisal Practices

Organizations vary on a wide array of factors (including size, product market, technology, culture, competitive environment, strategy, and union representation) that are likely to affect performance appraisal practices. Therefore, describing the modal setting in which appraisals take place is exceedingly difficult and should be approached cautiously. To reduce reliance on a single survey source, we thought it prudent to integrate the results of three recent surveys of current performance appraisal practices. These included a 1989 survey by the Wyatt Company of the performance appraisal practices in 3052 organizations (Wyatt, 1989), and a 1990 survey by The Conference Board of the performance appraisal practices in 435 of its member organizations (Milkoovich & Wigdor, 1991). Additionally, we conducted our own survey to assess the performance appraisal practices of the Fortune Industrial 100. The organizations represented in these three surveys varied in terms of size, location, and industry, and appeared to represent the diversity of U.S. private sector organizations. In all three surveys, the questionnaire was completed by a middle to high-level human resource manager.

For purposes of this review, the survey results are considered along four dimensions: (a) system design and characteristics, (b) system management, (c) important issues and current uses, and (d) performance distributions. Due to space limitations and because the samples differed in size and characteristics, we will present a qualitative integration of their findings. Readers interested in a more detailed presentation of the survey results should see Milkoovich and Wigdor (1991) and Wyatt (1989). Details regarding our Fortune 100 survey are available upon request.

System Design and Characteristics

Design. The performance appraisal systems in place in U.S. industry are on average 11 years old. They were designed primarily by personnel specialists with only limited input from the managers who use the system and virtually no input from employees affected by them. In light of the growing interest in employee and customer participation and involvement (Miller & Monge, 1986; Schweiger & Leana, 1986; Wagner & Gooding, 1989), customer service-oriented companies might rely on more customer-oriented companies. However, recently implemented systems are more likely to involve greater involvement in their design than were previous systems. 

Formats. Management-by-Objective in appraising executives, managers, and professionals, as well as trait-based rating scales are far less common. However, "mixed" formats are common, reflecting the multiple purposes for which appraisals are used. Two employee groups are far less likely to receive ratings: Executives are less likely to be formally appraised, while performance appraisals - reflecting a "top-down" appraisal process and negotiated agreements - are more likely in these cases.

Raters and sources of rating information. Ratings come directly from the immediate supervisor, with direct reports, peers, and subordinates, can provide valid additional input. However, the systematic integration of ratings of such sources is still very uncommon, as these ratings are typically filters through the immediate supervisor.

Quantitative indices are used to supplement rating information. Profits, sales, and costs are used to assess performance at all levels.

System Management

Time spent. It is common to spend 20 to 40 hours per year for each employee at lower levels and 80 to 100 hours per year for each employee at lower levels, although there is a wide variation on this issue. Many organizations reported spending between 20 and 40 hours per employee, whereas a few organizations spend considerably more, reporting spending between 20 and 40 hours per employee.

Decision making. Performance appraisal systems are often designed to foster a "top-down" approach, with the immediate supervisor making the final decision on an employee's performance. However, recent trends in the use of performance appraisals have emphasized greater input from subordinates, peers, and other stakeholders.
for the types of scales and formats employed have received considerable attention but the predominance of studies examined informally, yet virtually no systematic research affects the rater/ratee relationship or the way in which context affects employee perceptions of appraisal outcomes, and how appraisal performance moderates these relationships. Practicais in which appraisals are used in the current state of performance appraisal practices factors (including size, product market, and, strategy, and union representation) of appraisal practices. Therefore, describing the state of the art is exceedingly difficult and should be done on a single survey source, we thought it best to survey the current performance appraisal practices by the Wyatt Company of the performance appraisals practices in and 1990 survey of performance appraisal practices in 435 of its clients. Additionally, we conducted the appraisal practices of the Fortune 500 in these three surveys varied in terms of the diversity of U.S. private companies, the questionnaire was completed by a large number of respondents. The results are considered along four dimensions: (a) system management, (c) importance of performance distributions. Due to space limitations in size and characteristics, we will provide details. Readers interested in a more detailed treatment see Milkovich and Wigdor (1991). The Fortune 100 survey is also available upon request.

In U.S. industry are on average managed by personnel specialists with the system and virtually no input of the growing interest in employee development. As stated by Miller & Monge, 1986; Schweiger & Leana, 1986; Wagner & Gooding, 1987), we anticipated that more recently implemented systems might rely on more input from line managers, employees, and customers. However, recently implemented systems were more likely to have them involved in their design than were older systems.

**Formats.** Management-by-Objectives (MBO) is the preferred format for assessing executives, managers, and professional employees. Other formats, such as trait-based rating scales are far less common among these employee groups. However, "mixed" formats are common, and rating scales or ranking procedures are often used to supplement MBO-based approaches. In contrast, MBO is used to a far lesser extent among non-exempt employees (as defined by the Fair Labor Standards Act). For these employees, trait-based rating scales are the norm. A sizable minority of organizations report using behaviorally-based formats. However, pure behaviorally anchored rating scales (BARS), forced-choice scales, or mixed standard scales are very uncommon. Again, the use of "mixed" formats is common, reflecting the multiple purposes that appraisals serve in many organizations. Two employee groups are far less likely than others to receive formal evaluations. Executives are less likely to be formally evaluated — perhaps reflecting the difficulty of assessing performance at this level, or the reluctance of executives to submit to the process. Additionally, many hourly employees do not receive formal performance appraisals — reflecting, in part, organized labor's distrust of the appraisal process and negotiated agreements limiting the use of formal appraisals.

**Raters and sources of rating information.** The vast majority of performance ratings come directly from the immediate manager. For managerial and professional employees, the second level manager also has significant input. Recent research has demonstrated that non-traditional rating sources, such as self, peers, and subordinates, can provide valid appraisals. Moreover, the popular press has attested to the increasing use of such sources. However, it appears that these types of ratings are still very uncommon, and when they are used the information typically filters through the immediate manager who uses it in making his/her appraisal.

Quantitative indices are used to supply some performance information in most organizations. Profits, sales, and costs were frequently cited as important measures for executives and managers. The acquisition and use of job-specific knowledge was considered important for professional positions. Finally, attendance, quality, and quantity of output were important measures for nonexempt and hourly employees.

**System Management**

*Time spent.* It is common to spend about 7 hours per year assessing the performance of each employee at higher organizational levels and about 3 hours per year for each employee at lower levels. However, there is considerable variability on this issue. Many organizations report spending less than 1 hour per appraisal, whereas a few organizations spend considerably more time. For example, one reported spending between 20 and 40 hours per employee per year on the appraisal process.

*Decision making.* Performance appraisal policy decisions (e.g., whether to con-
duct formal appraisals, whether to link pay to performance) tend to be made at the corporate level in most organizations, but they are likely to be made at the business unit level in decentralized organizations. Decisions regarding appraisal practices (e.g., type of format to use, rater training issues), however, are as likely to be made at the business-unit level as they are at the corporate level. Very few organizations allow decisions about performance appraisal policies or practice to be made at the facility level.

**Training.** Most organizations report extensive use of rater training programs. However, training is most likely to occur when new systems are introduced, and few organizations provide rater training on an on-going (yearly) basis. Rater training is most likely to focus on conducting appraisal interviews and providing feedback, proper use of the new forms, setting performance standards, recognizing good performance, and avoiding rating errors. It seems, therefore, that performance appraisal practice has benefited from previous rater training research. However, raters receive virtually no training in how to best use the process to receive feedback or improve performance. Training remains focused on the rater (manager); preparing employees for their role in the appraisal process simply does not occur.

**Rater accountability.** It remains uncommon for managers to be evaluated on how they manage the appraisal process. Basic motivation models suggest that people will tend to behave in ways that maximize their expected payoffs or in ways for which they are reinforced (e.g., Vroom, 1964). In spite of this, only about one-quarter of the organizations surveyed attempted to hold raters accountable for how they managed the appraisal process. Among those that did, the most common method for doing so was to include it as a dimension on the rater’s own appraisal form.

**Important Performance Appraisal Issues and Uses**

Not surprisingly, managers consider fairness and justice issues to be very important. Most organizations report having an informal dispute resolution system (e.g., open door policies) that employees may use to contest the appraisal outcome. About one-quarter report having formalized processes available for this purpose (e.g., binding decisions made by a third party). However, a sizable minority reported that no appeals process was available. Although it may be common to have a mechanism for handling appeals, it is far less common to solicit employee opinions about the appraisal process. Most organizations do not systematically collect data to determine either the managers’ or the employees’ perceptions of fairness of the appraisal process or the results obtained.

Managers identified fairness as the most important performance appraisal issue organizations face. They also tend to be very concerned that the appraisal system be an effective tool to manage future performance, not just one that reflects past performance. Managers indicated that they are most likely to use performance information for improving future performance, making pay distribution decisions, and communicating expectations regarding future performance.

**Performance Distributions**

Performance appraisal systems typically rate employee performance. However, even though with five levels, generally only three levels are used with five levels, generally only three levels are used. Very few organizations expect large percentages of employees to receive low ratings and few, if any, to receive the low rating. It is common for 60% to be rated in the top two performance levels, placing performance on the part of the organization with leniency bias. Because the performance distribution is likely to be right-skewed, there is a common tendency for lower performance ratings to be given to more able and understanding employees, the distributions themselves evidence and researchers’ concern about the norm in U.S. industry is to rate employees’ performance distributions not only exist.

**The Juxtaposition of Rating Scales**

**Cognitive Processing Issues**

Research has indicated that expectations and prior expectations become an important part of the rating process. Because these expectations are based on knowledge or beliefs, raters will virtually always have prior expectations. It also seems that the type of appraisal system to which prior expectations became a part of the appraisal process is likely to persist over time. Raters, if they exist, they are salient key features in the appraisal system. Therefore, performance appraisal can be seen as an example of an apples and oranges type of situation. Additionally, the impact of prior knowledge and expectations created by different appraisal systems should not be overlooked.

The conditions under which actual performance reporting do not appear to be high, and with the realization that many organizations are pursuing the performance of employees in their work records, completing forms, preparing feedback. This is a relatively small amount, particularly when the performance of satisfactory performance is reported. Clearly demarcating the limits among different systems would be a good idea.
Performance Distributions

Performance appraisal systems typically have five levels to differentiate employee performance. However, even though most organizations report systems with five levels, generally only three levels are used. Both the desired and the actual distributions tend to be top heavy, with the top “buckets” relatively full and the bottom buckets relatively empty. Also, the actual distribution of performance is generally higher than the distribution desired. That is, even though most organizations expect large percentages of employees to receive the top performance ratings and few, if any, to receive the lowest ratings, their expectations tend to be conservative. It is common for 60 to 70% of an organization’s workforce to be rated in the top two performance levels. This could reflect either actual outstanding performance on the part of the organization’s workforce, or it could be indicative of leniency bias. Because the phenomenon is surprisingly constant across organizations, and it is unlikely that all organizations have predominately outstanding employees, the distributions probably reflect the latter. As anecdotal evidence and researchers’ concern about leniency have suggested, it appears that the norm in U.S. industry is to rate employees at the top end of the scale. Skewed performance distributions not only exist, but are common.

The Juxtaposition of Research and Practice

Cognitive Processing Issues

Research has indicated that expectations of future performance influence the rating process. Because these expectations are formed on the basis of prior knowledge or beliefs, raters will virtually always have some prior performance expectations. It also seems that the type of appraisal process used would affect the degree to which prior expectations became problematic. For example, in MBO-based systems, the mechanism through which a manager and subordinate arrive at mutually agreed upon goals requires that each individual form expectations regarding the level of performance that is achievable. Furthermore, because performance is measured against established goals, prior knowledge of job performance can also be expected. Therefore, prior expectations and prior knowledge not only exist, they are salient key features in the appraisal processes used in many organizations. Therefore, performance appraisal practice stands to benefit substantially from this line of research. Additionally, future research should explicitly consider the impact of prior knowledge and expectations under varying conditions of salience created by different appraisal systems.

The conditions under which actual appraisals occur also suggest that continued research on memory characteristics should prove to be valuable. This becomes apparent with the realization that managers report spending only a few hours per year assessing the performance of each employee. This time includes keeping records, completing forms, preparing for the appraisal interview and delivering feedback. This is a relatively small amount of time spread over a long period, particularly when the performance of several employees must be recalled and reported. Clearly demarcating the limits of memory and recall should lead to pro-
cesses, such as more frequent appraisals and systematic documentation, that will reduce reliance on memory.

Several other issues emerge from the cognitive processing literature. First, there has been a heavy reliance on student subjects and laboratory settings. Although there is some evidence that laboratory settings may provide results that are as valid and generalizable as those obtained in field settings (e.g., Locke, 1986), there is also convincing meta-analytic evidence that in the performance appraisal arena, effect sizes in paper-people studies are significantly larger than in studies involving observation of behavior (Murphy, Herr, Lockhart, & Maguire, 1986). Laboratory studies are often necessary in order to isolate particular effects. However, sterile environments that dilute the richness and complexity of the environment potentially change the phenomenon of interest. The potential effects of situational and contextual variables must be considered. The task of rating the performance of someone with whom an on-going relationship exists is both conceptually and operationally different from the rating task presented in laboratory settings. Therefore, though continued research on isolated cognitive processes is useful, research agendas should be expanded to include attempts to understand how these cognitive processes are affected by the political, social, and affective nature of most rating environments (Dipboye, 1985; Ferris & Judge, 1991).

Ratee/Ratee Personal Characteristics

The research on personal characteristics of raters and ratees is relatively balanced between laboratory and field settings and between student and employed samples. Given the labor markets trends expected to continue over the coming decade, research about age, gender, race, and ethnicity effects seems particularly timely and important. The labor force is expected to continue to age as the baby boom generation moves through. Additionally, the role of women and minorities is expected to increase, particularly among managerial ranks. Finally, the internationalization of the workforce introduces cultural differences regarding job design, performance expectations, and the role of performance feedback. Current research may be culture bound because it assumes a decidedly Western approach to these issues. It may be that employees raised in a traditional Japanese environment, for example, may expect a much less directive approach to appraisal and be offended by the confrontational nature of direct feedback. If individual difference characteristics such as these change the way ratings are assigned or interpreted, these trends represent significant issues to be addressed.

A potentially problematic issue that this literature needs to address is the tendency to find significant effects in studies using student samples but the absence of significant effects in field settings. This raises the question of whether the conditions encountered in experimental settings sufficiently capture the complexity of cross-gender relationships and sex-role stereotypes that exist in work settings.

Psychometric Issues

Researchers and managers appear to have different conceptualizations of accuracy. What does accuracy in performance appraisal imply? Many researchers would suggest that accurate appraisals are those that are both reliable and valid and conceptually near the true score. Managers tend to define accurate appraisals as those that allow the identification of relative differences within the context of the organizational environment in which it operates, involving deviations from true scores.

One potentially fruitful approach to this distinction between classification accuracy and behavioral accuracy (Burke & Judge, 1991) concluded that behavioral accuracy is more descriptive of what appraisals are supposed to mean, and may be more appropriate for which the ratings are made (for example, behavioral accuracy may be more important). Classification accuracy may be preferable (Murphy, 1991). Therefore, it appears that definition, and consequences of accuracy are important aspects of appraisal.

Appraisal Sources

It seems that an important point is to consider the sources. Rather than focusing on whether examining the psychometric properties of the appraisal, research should be examining the propriety of the conditions. That is, when should raters be asked to rate? Should they be asked to integrate with rating who does this? Should what this will become an even more important issue if the raters are different? For example, the increased use of peer appraisal and reduced hierarchical centered appraisals may also offer potentially specific situations. However, appraisal may be limited by the selection of raters (because only extreme cases are selected).

Suggestions for Design

First, surveys (including our own) in the practice, appear poorly suited to the respondents' characteristics. Most surveys are completed by a single person within an organization, and a single person knows enough about the raters of this type of complex human raters whether the views are expressed by the raters or by the organization. The very least, it would seem that multiple levels are necessary to accurately assess.

Second, the respondents to these th...
and systematic documentation, that will
cognitive processing literature. First, sit subjects and laboratory settings. Alter-
tory settings may provide results that are
field settings (e.g., Locke, 1986).
ence that in the performance appraisal
are significantly larger than in studies
order to isolate particular effects. How-
richness and complexity of the environ-
of interest. The potential effects of situ-
be considered. The task of rating the
ning relationship exists is both con-
the rating task presented in laboratory
search on isolated cognitive processes is
ed to include attempts to understand
ed by the political, social, and affective

cs of raters and ratees is relatively bal-
gs and between student and employed
expected to continue over the coming
and ethnicity effects seems particularly
expected to continue to age as the baby
ally, the role of women and minorities
 managerial ranks. Finally, the internal-
s cultural differences regarding job de-
directive approach to appraisal and be
direct feedback. If individual difference
way ratings are assigned or interpreted, 
be addressed.
 literature needs to address is the ten-
s using student samples but the absence
raises the question of whether the con-
gs sufficiently capture the complexity
stereotypes that exist in work settings.
we different conceptualizations of accu-
ance appraisal imply? Many researchers
rence that are both reliable and valid
and conceptually near the true score level of performance. However, managers
tend to define accurate appraisals as those that are accepted by employees and
allow the identification of relative contribution to organizational effectiveness
within the context of the organization and the constraints imposed by the regul-
atory environment in which it operates. This definition is quite different from one
involving deviations from true scores.
One potentially fruitful approach to resolving this disparity is Lord’s (1985)
distinction between classification and behavioral accuracy. Padgett and Ilgen
(1989) concluded that behavioral accuracy was both closer to the conceptual
criterion and more descriptive of what accuracy seems to mean in organizational set-
tings. However, because conveying behavioral information is seldom the sole
purpose of performance appraisal, Murphy (1991) points out that both types of
accuracy have meaning and may be differentially useful depending on the pur-
purpose for which the ratings are made (Murphy, Philbin, & Adams, 1989). For ex-
ample, behavioral accuracy may be more important when providing feedback, but
classification accuracy may be preferable for salary administration decisions
(Murphy, 1991). Therefore, it appears that additional research on the antecedents,
definition, and consequences of accuracy are in order.

Appraisal Sources

It seems that an important point is frequently overlooked in research on rating
sources. Rather than focusing on who should rate the performance of others and
examining the psychometric properties of various rating sources, perhaps re-
search should be examining the propriety of various rating sources under various
conditions. That is, when should ratings from alternative sources be used and how
should they be integrated with ratings from the immediate manager? It appears
that this will become an even more important issue as the nature of work con-
tinues to change. For example, the increasing use of self-managed work teams may
legitimize peer appraisals and reduce reliance on supervisory ratings. Customer-
centered appraisals may also offer potential for understanding performance in
specific situations. However, appraisals by customers may be problematic be-
cause customers may have limited job knowledge and these ratings are prone to
selection bias (because only extreme information is likely to be conveyed).

Suggestions for Directing Future Research

First, surveys (including our own) that have attempted to delineate the state of
practice, appear poorly suited to the task. The problems seem to be caused by re-
son respondent characteristics. Most survey methods send a single questionnaire to be
completed by a single person within each organization. It is questionable whether a
single person knows enough about the process to adequately convey the nu-
ances of this type of complex human resource system. It is also legitimate to ques-
tion whether the views expressed by a single organizational respondent are repre-
sentative of the organizational members for which he or she speaks. Therefore, at
the very least, it would seem that multiple perspectives, from multiple organiza-
tional levels are necessary to accurately describe the organization’s practices.

Second, the respondents to these types of surveys are typically middle or high-
level human resource managers that have some kind of policy-making role in the organization. This raises the question of whether their responses are descriptive of the appraisal system as practiced or as it was intended to be practiced. Given the vested interest the typical respondent has in the appraisal system being described, one might argue that it would be rational to paint as favorable a picture as possible. Therefore, we suggest that future survey research (a) use multiple respondents from each participating organization, and (b) clearly distinguish between how the appraisal process was intended to be used and how it is actually used.

The surveys of current practices raise some fundamental research issues. Performance appraisal systems may be considered to be a series of decisions that are affected by environmental, organizational, and dispositional factors. Research is needed to examine the variety of situational variables that affect appraisal design and administrative choices. For example, what are the determinants of managerial choices in performance appraisal design and administration? What factors lead managers to choose to decentralize policy and administrative responsibilities, to select a forced distribution approach, to use more objective performance indices, to place more weight on second level managers as sources of information, to change the number of levels in the appraisal scale or even to determine the desired performance rating distributions? These types of questions are illustrative of the myriad of decisions that go into implementing any performance appraisal system. Research has tended to focus on the outcomes of these decisions. However, we seem to know very little about the factors that cause decisions makers to implement certain approaches. Research directed at these types of issues would seem particularly useful for informing future practice.

The surveys revealed some appraisal decisions that varied across employee groups and others that did not. Specifically, though MBO is the most common approach for assessing managers, graphic rating scales are more common among nonexempt employees, and many executives are not subjected to any formal appraisal process. These differences are potentially problematic because the appraisal process appears to become less standardized and systematic at higher organizational levels. It seems that these differences represent potential research opportunities. For example, research could examine whether this practice affects perceptions of procedural and distributive justice both between and within employee groups. We suspect that it does, but research as to the behavioral consequences of these types of practices would prove insightful.

The surveys do not convey a sense of how organizations tie performance appraisals to their underlying culture. Differences in this regard may have implications for organizational effectiveness. Are there specific performance appraisal system characteristics that are better suited to particular climates or cultures? An obvious characteristic to consider is the degree to which performance ratings drive pay decisions. Specifically, stronger pay for performance contingencies might be more important under conditions of intense competition than in regulated or cost-plus environments. Also, perhaps administrative uses of performance data should be emphasized in a meritocracy, but developmental uses should be emphasized in public sector/civil servant environments. Finally, different types of organizations certainly reward different types of behavior. Olian and Rynes (1984) speculated, for example, that very different types of behavior are addressed these speculations. Might interactions are more willing to set "stretch goals" to establish voice mechanisms? Answers may create a taxonomy of "fit" between personnel and organizational environments. Such research would delineate more clearly the modular organizational settings.

Napier & Latham (1986) found that positive or negative, of conducting performance reviews, the use of a peer and a boss may vary. For example, one study found that contextual variables (such as part-time work) influence satisfaction (Dobbins, Cardy, & Platz). With these exceptions, recent research suggests that the appraisal system designs or processes may vary within an organization.

Some characteristics would appear to be constant; research also might address whether raters are willing to set "stretch goals" or would supposedly lose the ability to do so. On the other hand, too few levels are likely to be too few; conversely, what is too many or too few? Concerns in which the top and bottom levels are still problematic cases. The middle level is often the most problematic. If performance is normally distributed, the standard deviation below average would be higher than average, the standard deviation below average would be lower than average. The problem is that the distribution of performance is not always normal. Deming (1986) argues strongly that organizations that rely on rules and routines and manage by fact, rather than by authority, are likely to be more productive. This approach, reported in wide use at agricultural facilities, focuses on assisting those who are doing work and to control by fact, rather than by authority, is likely to be more productive. A similar approach, which emphasizes the informal feedback system that seems difficult to maintain, is used in informal appraisals by team leaders in essentially less systematic and more vulnerable environments.
Rynes (1984) speculated, for example, that prospectors and defenders would reward very different types of behavior, but we are not aware of research that has addressed these speculations. Might it also be that particular types of organizations are more willing to set “stretch goals,” forgive performances deficiencies, or establish voice mechanisms? Answers to these types of questions might begin to create a taxonomy of “fit” between performance appraisal system characteristics and organizational environments. Such a taxonomy would certainly allow research to delineate more clearly the most effective appraisal practices for particular organizational settings.

Napier & Latham (1986) found that managers perceived no consequences, positive or negative, of conducting performance appraisals. Conversely, Longenecker, Sims and Gioia (1987) reported that because of actual and perceived negative consequences of accurate appraisal, some managers knowingly make ratings that are inaccurate. However, the rating environment likely affects both rating practices and participant reactions. For example, recent research has shown that contextual variables (such as participation and rating frequency) affect ratee satisfaction (Dobbins, Cardy, & Platz-Vieno, 1990; Giles & Mossholder, 1990). With these exceptions, recent research has not examined the effects of different appraisal system designs or processes on employees’ attitudes and behaviors.

Some characteristics would appear to influence participant attitudes. For example research also might address whether the use of three, four, or five (or more) rating levels affect employee motivation or work attitudes. Too many levels would supposedly lose the ability to differentiate performance meaningfully. On the other hand, too few levels are likely to create perceptions of inequity. However, what is too many or too few? Consider the proto-typical three-bucket system in which the top and bottom levels are reserved for truly outstanding and truly problematic cases. The middle level is designed to capture 75 to 80 percent of the workforce. If performance is normally distributed, an employee performing one standard deviation below average would likely receive the same rating as an employee performing one standard deviation above average. The better performers would likely consider this to be unfair. What are the perceptions of employees at different positions in the distribution? Are the better performers in the middle bucket more likely to withdraw than are the poorer ones? Are they more likely to withdraw than are similar performers in say five-level systems in which their performance would be differentially acknowledged?

Deming (1986) argues strongly that performance appraisal has serious negative consequences and urges organizations to cease all individual performance appraisals and to evaluate unit or plant level performance instead (Scholtes, 1987). This approach, reportedly in wide use in high involvement - high commitment facilities, focuses on assisting those whose performance is “out of the system.” Deming’s notions have received some attention in practitioner performance appraisal literature, but no attention from researchers. The possibility of no individual feedback seems difficult to attain. Even without formal individual appraisals, informal appraisals by team leaders and peers seem inevitable and perhaps potentially less systematic and more vulnerable to biases. Accordingly, research might
address the effects of informal appraisal processes on employee perceptions, attitudes, and behavior.

The effects of skewed performance distributions on pay allocation and employee attitudes also needs to be examined. Highly skewed ratings affect the distribution of merit pay increases. In fact, some argue that the size of the merit fund pool affects the ratings skew (Milkovich & Newman, 1991). Smaller funds may force managers to give higher ratings, which result in smaller average increases for high rated performers. These smaller pay increases coupled with high ratings deliver mixed signals to employees and may affect their attitudes toward merit pay. In this regard, tracking changes in appraisal system decisions over time may be useful. For example, some firms report shifting to forced distributions in an effort to correct their ratings skew. Such changes may lower employees’ satisfaction with appraisals on the one hand, but improve employees’ satisfaction with their pay increases on the other. More generally, it seems desirable to understand why firms make changes in their systems and to examine the effects of these changes.

Conclusion

In concluding, we suggest that performance appraisal research and practice seem to converge on many issues and diverge on others. Divergence on some issues is not necessarily a problem because relevancy for decision makers is not the purpose for all research efforts. Yet performance evaluation is an applied subject and, as such, research should eventually lead to improvements in practice. Continued reliance on student samples and laboratory settings is not facilitating the transfer of research into application. We do, however, need better understanding of the information-processing capabilities and limitations of human decision making.

We also need to continue developing a more comprehensive theory of the rating process. Since Wherry’s work in the 1950s’ (see the Appendix to Landy & Farr, 1983), the collection of studies on information processing is the most serious, concentrated attempt to date to better understand the rating process. In that framework, continued research along those lines is useful indeed. However, attention must be paid to the potential effects of situational or contextual variables. Examining appraisal issues in sterile environments may isolate the effects many researchers wish to investigate, but also limits the generalizability of the results and removes the issues from the attention and interests of human resource decision makers. If research is to inform practice, interaction between researchers and managers and application of research results are important.

On the other hand, organizations continue to do things that undermine the effectiveness of the appraisal process. Little time is spent on the appraisal process; raters are not systematically trained and are not held accountable. The employee’s role in the performance process is overlooked as are many potentially valuable sources of performance information (self, peers, subordinates). Although research has done much to suggest improvements regarding many of the practices noted above, it may ultimately be the changing nature of work that leads managers to implement practices that research has legitimized. For example, the trend toward self-managed work teams is diminishing the relationship. On the one hand, this may provide for more fairer appraisals; on the other hand, it is likely that the cognitive processes involved in supervisory ratings due to task interactions resulting from peer relationships.

The issues of interest to managers and are mutually exclusive. Managers are concerned with systems that help them manage more success. Efforts to understand how information is processed may be useful. Assuming bias assumptions, limiting these factors may result in fairer appraisals. Therefore, managerial concerns by cognitive-processing research. New roles for the appraisers are receiving little or no attention. This is the need for a more explicit focus on the presence of a wide variety of unique opportunities to examine the determinants of behavior in organizations. Although a considerable body of theory is most interestingly under conditions where performance increases or performance ratings are judged.

Therefore, it appears that current performance evaluation is expanded to include these concerns. Although it is necessary, an expanded research agenda that would further allow performance appraisal theory to influence organizations.

References


Processes on employee perceptions, attributions on pay allocation and employee performance. Highly skewed ratings affect the discipline and the size of the merit fund (Newman, 1991). Smaller funds may result in smaller average increases by increases coupled with high ratings may affect their attitudes toward merit system decisions (e.g., lower employees' satisfaction with improve employees' satisfaction with rating, it seems desirable to understand and to examine the effects of these

Appraisal research and practice are on others. Divergence on some issues for decision makers is not the same evaluation is an applied subject to improvements in practice. Conventional settings is not facilitating the important, need better understanding and limitations of human decision making. The importance of understanding the generalizability of the results and interests of human resource decision interaction between researchers and practitioners. To do things that undermine the effectiveness is spent on the appraisal process; not held accountable. The employee’s (as many potentially valuable workers, subordinates). Although research regarding many of the same nature of work that leads managers to mixed. For example, the trend toward

Self-managed work teams is diminishing the traditional supervisor-subordinate relationship. On the one hand, this may ultimately lead to greater acceptance of peer appraisals; on the other hand, it is likely to force research into new directions as well. The cognitive processes involved in peer appraisals are likely to differ from those in supervisory ratings due to the differences in power and social interactions resulting from peer relationships.

The issues of interest to managers and researchers may be different but not mutually exclusive. Managers are concerned with fairness and using appraisal systems that help them manage more effectively. Cognitive-processing research attempts to understand how information is translated into ratings so that bias and error may be removed. Assuming bias and error contribute to suboptimal decisions, limiting these factors may result in better decision making and ultimately fairer appraisals. Therefore, managerial concerns for fairness are being addressed by cognitive-processing research. Nevertheless, some very important issues raised by managers are receiving little or no research attention. Most important of these is the need for a more explicit focus on procedural and distributive justice. Although a considerable body of theoretical discussion exists, appraisals offer unique opportunities to examine the determinants of fair procedures under varying conditions (e.g., different occupational groups, across firms), and perhaps most interestingly under conditions when the distributive results, such as pay increases or performance ratings, are judged to be unfair (Greenberg, 1988, 1990).

Therefore, it appears that current performance appraisal research could be expanded to include these concerns. Although the current focus is productive and necessary, an expanded research agenda that included the issues discussed above would further allow performance appraisal research to influence human behavior in organizations.

References


---


of the relationship between the organizational

discussion: Tacit influences on performance app-
ance: An initial investigation of the joint effects of organizational variables and appraisal char-

diables: A practical guide. Public Personnel

appraisals: Enhancing managerial productivity.

appraisals and noncompensatory processes in performance

standards: A practical guide. Public Personnel

appraisals: Liking as an integral dimension of

conquering evaluation fear. Personnel Adminis-

appraisal in higher education: Comparing de-

1989. Multiple uses of performance appraisal:

joins. Personnel Journal, September: 48-55,

performance appraisal tool. Public Personnel

notional disability and clear performance stan-

and performance appraisal: A research note on

Personnel, August: 43-45.

Harvard Business Review, September-October

proved its performance appraisals. Personnel

performance feedback: Reactions to alternate

s. A cognitive model of the performance ap-

s. Organizational Behavior and Human Perfor-

DeNisi, A. S., Robbins, T., & Cafferty, T. P. 1989. Organization of information used for perform-


Dickinson, T. L. 1987. Designs for evaluating the validity and accuracy of performance ratings. Organiza-


Dobbins, G. H., Cardy, R. L., & Platz-Vieno, S. J. 1990. A contingency approach to appraisal satisfac-

Dobbins, G. H., Cardy, R. L., & Truxillo, D. M. 1988. The effects of purpose of appraisal and indi-


Dugan, K. W. 1989. Ability and effort attributions: Do they affect how managers communicate perfor-


Farh, J. L., & Dobbins, G. H. 1989. Effects of comparative performance information on the accu-


Fedor, D. B., & Rowland, K. M. 1989. Investigating supervisor attributions of subordinate perfor-


JOURNAL OF MANAGEMENT, VOL. 18, NO. 2, 1992

JOURNAL OF MANAGEMENT, VOL. 18, NO. 2, 1992


Nathan, B. R., & Tippins, N. 1990. The consequences of halo "error" in performance ratings: A


Vance, R. J., MacCallum, R. C., Coover, M. D., & Hodge, J. W. 1988. Construct validity of multi-


APPENDIX

Summary of Research and Managerial Performance Appraisal Literature

<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal: A neglected aspect of the performance appraisal</td>
<td>Landry &amp; J.I. Farr</td>
<td>1990</td>
</tr>
<tr>
<td>Human Decision Processes</td>
<td>T. P. Cafferty</td>
<td>1986</td>
</tr>
<tr>
<td>Initial decisions and sub-influence in performance appraisal</td>
<td>T. P. Cafferty</td>
<td>1986</td>
</tr>
<tr>
<td>Review approach to improving productivity</td>
<td>Per-Asian survey</td>
<td></td>
</tr>
<tr>
<td>Influence in performance appraisal</td>
<td>J. A. Cafferty</td>
<td>1984</td>
</tr>
<tr>
<td>Participation-outcome relationships</td>
<td>Academy</td>
<td>1990</td>
</tr>
<tr>
<td>HR Magazine, July: 66-90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance by appraisal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference Executive, May: 36-38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personnel, February: 42-46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factors affecting and organizational behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis of performance appraisal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal of Applied Psychology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance appraisal: An update</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal of Applied Psychology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary of Research and Managerial Performance Appraisal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literature Review</td>
<td>Methodology Survey</td>
<td>Discussion</td>
</tr>
<tr>
<td>Lab Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Issues Explored</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per-Asian Survey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Decisions and Subshows</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review Approach to Improving Productivity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Influence in Performance Appraisal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

JOURNAL OF MANAGEMENT, VOL. 18, NO. 2, 1992
<table>
<thead>
<tr>
<th>Primary Issues Explored</th>
<th>Field Study</th>
<th>Lab Study</th>
<th>Lit Review</th>
<th>Methodology</th>
<th>Survey</th>
<th>Discussion</th>
<th>Case Study</th>
<th>&quot;How To&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kusso &amp; DeNisi, 1989</td>
<td>Karl &amp; Wesley, 1989</td>
<td></td>
<td>Focareo, 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Murphy et al., 1989</td>
<td></td>
<td>Sankowsky, 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pagett &amp; Ilgen, 1989</td>
<td></td>
<td>Sautter, Barry, &amp; Reilly, 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sautter, Collins, &amp; Bude, 1989</td>
<td></td>
<td>Sautter, 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dortman et al., 1986</td>
<td>Dugan, 1989</td>
<td></td>
<td>Wight, 1985</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Earley, 1988</td>
<td>Fedor et al., 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Finb &amp; Porter, 1986</td>
<td>Hearn &amp; Hearn, 1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prince &amp; Lawler, 1986</td>
<td>Kinowski &amp; Inks, 1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Russell &amp; Goode, 1988</td>
<td>Northcraft &amp; Ashford, 1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raters/Appraisal Sources</td>
<td>Eder &amp; Fehr, 1989</td>
<td>Fath &amp; Dobson, 1989</td>
<td>Harris &amp; Schauberger, 1988</td>
<td>Bernardin, 1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Skopen &amp; Dieder, 1985</td>
<td></td>
<td></td>
<td>Sautter, 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vance et al., 1988</td>
<td></td>
<td></td>
<td>Sautter, 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formats</td>
<td>Harris, 1988</td>
<td></td>
<td></td>
<td>Martin &amp; Bartol, 1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tenopyr, 1988</td>
<td></td>
<td>Oliver, 1985</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primary Issues Explored</th>
<th>Field Study</th>
<th>Lit Review</th>
<th>Methodology</th>
<th>Survey</th>
<th>Discussion</th>
<th>Case Study</th>
<th>&quot;How To&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Napier &amp; Latham, 1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pooyan &amp; Eberhardt, 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal Rationalizations</td>
<td>Rippin &amp; Schmidt, 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Longenecker et al., 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wayne &amp; Ferris, 1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wesley &amp; Snell, 1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social and Political Aspects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Krippin &amp; Schmidt, 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Longenecker et al., 1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wayne &amp; Ferris, 1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wesley &amp; Snell, 1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rates/Appraisal Sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fels &amp; Feder, 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fells et al., 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fox &amp; Hume, 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fox et al., 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>George &amp; Smith, 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>McAlvey &amp; Butler, 1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shapiro &amp; Dozier, 1985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vance et al., 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rater Training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duggan, 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hodge &amp; Kavanagh, 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Athey &amp; McAlvey, 1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Smith, 1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Banks &amp; Robertson, 1985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weltensberg, 1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Harris, 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hughes &amp; Prien, 1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prien &amp; Hughes, 1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Murphy &amp; Consants, 1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Murphy &amp; Pandolff, 1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Terrey &amp; Bowers, 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primary Issues Explored</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field Study</td>
</tr>
<tr>
<td>Lab Study</td>
</tr>
<tr>
<td>Methodology</td>
</tr>
<tr>
<td>Lab Review</td>
</tr>
<tr>
<td>Survey</td>
</tr>
<tr>
<td>Discussion</td>
</tr>
<tr>
<td>Case Study</td>
</tr>
<tr>
<td>&quot;How To&quot;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reactions To Appraisals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dobson et al., 1990</td>
</tr>
<tr>
<td>Giles &amp; Mesholmer, 1990</td>
</tr>
<tr>
<td>Naylor &amp; Luftman, 1988</td>
</tr>
<tr>
<td>Pooyan &amp; Eberhard, 1989</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Legal Ramifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrett &amp; Kerness, 1987</td>
</tr>
<tr>
<td>Griffin, 1989</td>
</tr>
<tr>
<td>Miller et al., 1990</td>
</tr>
<tr>
<td>Daly, 1990</td>
</tr>
<tr>
<td>Goddard, 1989</td>
</tr>
<tr>
<td>Metz, 1988</td>
</tr>
<tr>
<td>Vincent, 1990</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social and Political Aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kipnis &amp; Schmidt, 1986</td>
</tr>
<tr>
<td>Lengenacker et al., 1987</td>
</tr>
<tr>
<td>Wayne &amp; Harris, 1990</td>
</tr>
<tr>
<td>Wexley &amp; Scull, 1987</td>
</tr>
<tr>
<td>Brimberg, 1985</td>
</tr>
<tr>
<td>Kerr &amp; Joyce, 1989</td>
</tr>
<tr>
<td>Perry &amp; Parnell, 1988</td>
</tr>
<tr>
<td>Schneier, 1989</td>
</tr>
<tr>
<td>Day, 1989</td>
</tr>
<tr>
<td>Kirkpatrick, 1986</td>
</tr>
<tr>
<td>Levy, 1989</td>
</tr>
<tr>
<td>Romans/II, 1989</td>
</tr>
<tr>
<td>Schneier et al., 1986</td>
</tr>
<tr>
<td>Schneier et al., 1989</td>
</tr>
<tr>
<td>Schneier et al., 1987</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland et al., 1980</td>
</tr>
<tr>
<td>Locher &amp; Teal, 1988</td>
</tr>
<tr>
<td>Warrenlitz, 1989</td>
</tr>
<tr>
<td>Wyatt, 1989</td>
</tr>
<tr>
<td>Devens, 1990</td>
</tr>
<tr>
<td>Edwards, 1990a</td>
</tr>
<tr>
<td>Longnecker &amp; Gora, 1988</td>
</tr>
<tr>
<td>Bayer et al., 1988</td>
</tr>
<tr>
<td>Cocher, 1986</td>
</tr>
<tr>
<td>Cozzens, 1990</td>
</tr>
<tr>
<td>Diers &amp; Tyler, 1996</td>
</tr>
<tr>
<td>Cutlerman &amp; Hodgson, 1989</td>
</tr>
<tr>
<td>Fladd, 1988</td>
</tr>
<tr>
<td>Goodell, 1988</td>
</tr>
<tr>
<td>Gomes-Mejia et al., 1985</td>
</tr>
<tr>
<td>Hall, 1987</td>
</tr>
<tr>
<td>Laeunemeyer &amp; Bebe, 1988</td>
</tr>
<tr>
<td>Scherkenbach, 1985</td>
</tr>
<tr>
<td>Wagel, 1987</td>
</tr>
<tr>
<td>Woods &amp; Wilcoxon, 1985</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PA Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clemmer &amp; Stevens, 1989</td>
</tr>
<tr>
<td>Rentz, 1985</td>
</tr>
<tr>
<td>Cleveland et al., 1980</td>
</tr>
<tr>
<td>Locher &amp; Teal, 1988</td>
</tr>
<tr>
<td>Warrenlitz, 1989</td>
</tr>
<tr>
<td>Wyatt, 1989</td>
</tr>
<tr>
<td>Devens, 1990</td>
</tr>
<tr>
<td>Edwards, 1990a</td>
</tr>
<tr>
<td>Longnecker &amp; Gora, 1988</td>
</tr>
<tr>
<td>Bayer et al., 1988</td>
</tr>
<tr>
<td>Cocher, 1986</td>
</tr>
<tr>
<td>Cozzens, 1990</td>
</tr>
<tr>
<td>Diers &amp; Tyler, 1996</td>
</tr>
<tr>
<td>Cutlerman &amp; Hodgson, 1989</td>
</tr>
<tr>
<td>Fladd, 1988</td>
</tr>
<tr>
<td>Goodell, 1988</td>
</tr>
<tr>
<td>Gomes-Mejia et al., 1985</td>
</tr>
<tr>
<td>Hall, 1987</td>
</tr>
<tr>
<td>Laeunemeyer &amp; Bebe, 1988</td>
</tr>
<tr>
<td>Scherkenbach, 1985</td>
</tr>
<tr>
<td>Wagel, 1987</td>
</tr>
<tr>
<td>Woods &amp; Wilcoxon, 1985</td>
</tr>
</tbody>
</table>
Organizational Stress and Job Performance: Where Do We Go From Here?

Sherry E.,
Memphis State University

This article reviews and summarizes the literature concerned with both direct and indirect analyses of the relationship of organizational stress and job performance. Moderate methodological operationalizations at the individual level of analysis are classified and then reviewed. A synthesis and summary of this research suggests significant improvements in the analysis of such phenomena, much of this research needs to focus on the study of reciprocal relationships that are the benchmark for guidelines for improving the quality of organizational inquiry.

Despite recent advances in research, there is considerable disagreement on basic theoretical and methodological assumptions about the nature of the relationship between organizational stress and job performance, and the empirical evidence on these relationships. This article presents a critical review of the literature on organizational stress and job performance, with a focus on the methodological and theoretical issues that need to be addressed in future research.

The preparation of this article was supported in part by the Graduate School of the University of Memphis. Address all correspondence to Dr. Sherry E. Salas, Department of Psychology, Memphis State University, Memphis, TN 38152.

Copyright 1992 by the Southern Management Association.