

Performance Appraisal: Behavioralism and Its Discontents

Charles A. Schwartz

In the area of performance appraisal, libraries and other nonprofit organizations face a special dilemma of providing job feedback while maintaining a nonauthoritarian climate. This paper assesses the utility of four approaches to this problem: behavioralism, minimal model, heuristic literature, and tacit knowledge from in-house experimentation. Part one argues that sources of rater error are generally intractable and, concomitantly, that behavioral appraisal methods are ineffective. Part two takes up the widespread notion that performance appraisal, as a matter of ethics, should be less ambitious and important. The third part describes the heuristic literature in this field as a "folk-science." The next part surveys the experience with in-house experiments in the social and behavioral sciences. Then, the final part of the paper draws some conclusions about the relative utility of the different approaches and offers a suggestion as to the future procedure in this field.



he literature on performance appraisal has been written largely in terms of two contrasting emphases. On the one hand are those writers, principally in the fields of business and applied psychology, who propose multitrait-multimethod appraisal systems. This behavioral approach has a rich, half-century history of research, but it has developed unevenly. Investigation of most of the possible types of rating formats has probably been pushed to the limits of its usefulness, while study of organizational factors surrounding the appraisal process is still rudimentary.¹ Underlying the concepts and procedures of the behavioral approach is the basic notion that some combination of appraisal methods can be devised to control rater error.

At the other extreme is a smaller group, which argues that appraisals should be scaled down and made less important, es-

pecially for salary and rewards administration. This argument is roughly a quarter-century old but has not been advanced empirically. It draws on Douglas McGregor's classic contention that supervisors and subordinates distrust appraisals for fundamentally sound reasons: few supervisors want to play God; at the same time, every subordinate seeks to protect the integrity of his or her own personality, as well as to obtain valued rewards.²

In the gap between the two emphases is a heuristic perspective common to journals geared toward practitioners (e.g., *Harvard Business Review*) and to library science. It is found mainly in lists of recommendations about what to avoid when evaluating people and how to motivate them in the process.

Although the literature in the library field over the past ten years testifies to a growing concern about the need for better

Charles A. Schwartz is Social Sciences Bibliographer at the University of Northern Iowa, Cedar Falls, Iowa 50613.

appraisal methods,³ general discussions have not been followed up by experimental research that describes how particular methods affect performance at work, the state of morale, or the margin of error.

As an approach to performance appraisal, in-house experimentation is a novelty in the social and behavioral sciences. In part, interest in it can be traced to the gradual unfolding of two interdisciplinary concerns that pervade the so-called post-behavioral movement. These concerns are the issue of certainty versus usefulness of research and the question of whether human interaction and social arrangements can be understood with a disengaged consciousness. However, the actual practice of in-house experimentation has been limited because of the difficulty involved.

This paper assesses the utility of different approaches to performance appraisal for nonprofit organizations. The first section reviews the behavioral approach in three areas: (1) sources of rater bias and variance in the traditional supervisor-subordinate context; (2) problems inherent in alternative appraisal contexts; and (3) the status of several common appraisal techniques. The second section discusses the minimal appraisal, with particular reference to the dilemma of providing better job feedback while preserving a non-authoritarian climate. The third section describes the heuristic approach to performance appraisal, focusing on two problems in this area: the valuing aspect of appraisals, and the limitations of common sense. The fourth section contains a review of the experience with in-house experiments. The fifth and concluding part provides a brief summary that calls for adaptation of in-house experimentation and development of tacit (personal) knowledge.

The term *performance appraisal* must be broadly interpreted since it covers placement and promotion decisions, salary and rewards administration, motivational and developmental counseling, and criteria for the evaluation of personnel practices. This paper discusses all of these matters, but focuses on the last one; legal concerns, such as due process, are not discussed.

The term *nonprofit organization* requires comment. Administration texts distinguish among three sectors: business enterprises, government agencies, and nonprofit organizations. Nonprofit organizations are distinguished from the other sectors by having ambiguous means and ends, little or no job feedback, and a tendency not to monitor their environments. Whereas the characteristics and earnings of nonprofit workers may differ substantially from workers in business or in government, there is no evidence that job orientations or other subjective behaviors vary across the sectors.⁴ But the term is preferable to some others, such as *knowledge-based industries*⁵ or *organized anarchies*.⁶

I. BEHAVIORALISM

Rater Bias and Variance

There exist severe problems in obtaining (1) accurate supervisor ratings of job behavior, even on a single performance dimension; and (2) agreement among raters on a ratee's composite, comparative evaluation. Although rater agreement is no guarantee of accurate results, rater disagreement does signal the presence of error. Two fundamental sources of rater variance can be singled out: bias and halo error. Many specific sources of variance can be partially remedied or at least accounted for in appraisals, whereas bias and halo error remain intractable.⁷

By all accounts, rater bias (i.e., preference or prejudice, whether conscious or unconscious) is the most serious common problem of performance appraisals. A substantial body of empirical literature demonstrates what one might expect from everyday experience: people tend to rate people like themselves more favorably.⁸ According to personal construct theory, each individual relies on a set of his or her own characteristics to make judgments about others. Thus, "more effective" managers value initiative, planning ability, perseverance, and broad knowledge; "less effective" managers value consideration, tact, cooperation, teamwork, and loyalty.⁹ In the same way, raters are differentially accurate in identifying "correct" and "incorrect" job behaviors.¹⁰ Even

when raters focus on a single behavior or trait, subjectivity breeds variance.¹¹

Halo error is committed by raters relying on a summary impression of each ratee rather than distinguishing among the levels of performance that individuals exhibit on different job dimensions. Evaluation of any one dimension is not synonymous with performance appraisal as a whole.¹² A number of reports conclude that no appraisal instrument or amount of education or experience is satisfactorily resistant to halo error.¹³ Training sessions in which participants are urged not to commit this error run two related risks. One is that individual raters are not capable of accurately identifying ratees' strong and weak performance areas. "Thus, it is possible that reducing halo may also lead to reduced reliability."¹⁴ The other risk is that "making greater distinctions among different performance categories may lead to a reduction in interrater reliability" as well.¹⁵

Possible upper limits in reliability of performance ratings were explored in one experiment by creating a nearly ideal environment in which actors played jobs familiar to the raters. Results indicated an inborn "ceiling for interrater agreement in judging performance."¹⁶ Given what appear to be human limitations associated with halo error, some behavioral researchers are turning away from efforts to control halo error and toward study of the kinds of cues people depend upon in making summary evaluations of others. Personal construct theory is one example of this shift in theoretical approach.

Another example of the new research centers on attribution theory in the field of social psychology.¹⁷ This theory seeks to explain by a two-stage process people's naive assumptions about the causes of their own and others' behavior. In the first or diagnostic stage, the supervisor, presented with an incident of poor performance (e.g., tardiness, a missed deadline), tries to determine the cause by sorting through a variety of informational cues. Such attribution involves a judgment about whether the cause was internal to the subordinate (e.g., his or her attitude, ability, or effort) or external to the

subordinate (e.g., task difficulty, lack of support, insufficient information).

The diagnostic stage is typically dominated by self-serving biases. A number of empirical studies show that people think their own behavior stems from external forces, while others' behavior stems from internal factors.¹⁸ Thus, in the case of unsatisfactory job performance, the supervisor is likely to blame internal causes (i.e., the ratee), while the subordinate is likely to blame external events (i.e., things beyond his or her control).

At the decision stage the supervisor can either modify the job or tell the subordinate to "change." Usually, the latter course is easier. This tendency toward internal attribution is strengthened when the outcome of unsatisfactory performance is serious—even though the outcome may have been beyond the subordinate's control. However, when poor performance occurs without serious consequence, the supervisor is more likely to overlook the problem.¹⁹ Attribution theory holds that, for more effective appraisals, supervisors should focus on the behavior and not the outcome.

Alternative Rating Patterns

In view of the inherent difficulties of the supervisor-subordinate relationship, various other rating patterns have received attention. Research on subordinate evaluation of the supervisor is slim; also, the results of experiments do not lend themselves to clear generalization, apart from the evidence of a potential for frictions and misunderstandings in the use of this appraisal.²⁰

Laboratory experiments, lacking any real threat of ratee retaliation, are theoretically interesting but tend to be artificial. One set of experiments on reverse evaluation suggests a declining role for sex bias. Traditionally, women have been rated differently from men, very often more negatively.²¹ At the University of Dayton, male subjects and female confederates worked in dyads on a task in which the female confederate was always chosen leader over the male subject. The selection of leaders was designed to appear in one of three ways to the men: as an arbitrary decision

based on sex; as a matter of "chance"; or as a determination of "merit," based on a (secretly rigged) male-female competition on a prior task. It was arranged for half of the teams to succeed at the task, the other half to fail.²²

Results of the experiments indicate that appraisal of women as leaders is mainly a dual function of perceived fairness of the selection method and of task outcome. In the first instance, male subordinates who think that a woman was handed a leadership position by virtue of her sex may try to restore psychological equity by downgrading a woman's qualifications. The other, more powerful criterion of appraisal for the woman leader is whether she succeeds at the task.

Although these findings can reasonably be extended to a leader of either sex when the selection process is perceived to be unfair, they carry a particular warning against clumsy affirmative action programs. Also, they suggest that organizations can best insure a person's acceptance in a leadership role by assigning relatively easy tasks at first.

Peer evaluations have a curious quality. They tend to show more halo error than supervisor ratings (i.e., less sensitivity to differences among job behaviors);²³ yet, they have high reliability in the prediction of leadership potential.²⁴ No general explanation of this reliability seems to exist, but a number of specific findings are well documented: (1) friendship plays a minor role; (2) racial bias is to be expected; (3) brief observations of physical appearance and manner are highly similar to observations based on considerable prior contact; and (4) peer evaluations are stable across groups, or as group composition changes.²⁵

Self-ratings are dichotomous. When employees compare their own job performance with the that of others in similar jobs at about the same salary, self-ratings are greatly inflated. As a rule, the higher the job level, the greater the inflation. For example, among managers more than 80 percent rated themselves in the top 10 percent category.²⁶

On the other hand, self-appraisals for developmental purposes are much more reliable indicators of one's relative

strength in different capacities. Research has shown that self-appraisals of managerial potential, when done confidentially, have fairly good job-behavior discriminant validity and are not unduly influenced by halo error.²⁷ Overall, most people have an intuitive grasp of their ability despite an unrealistically favorable perception of their current performance compared to others' performance. This self-delusion is not necessarily bad. As research has also shown, high self-esteem is associated with many desirable work characteristics, such as motivation, pride in one's work, and physical and mental well-being.²⁸

Appraisal Techniques

Rater Training. Considerable attention has been directed toward attempts to reduce various rating errors through rater training. Recent studies suggest few solid leads about how to train people to observe and record job behaviors reliably.²⁹ Some writers argue that training programs should not be encouraged because, lacking any theoretical model, "we do not know what truth in performance appraisal is."³⁰ Others contend that some training, at least in communication skills, is necessary because raters without any training "have the potential of doing more harm than good."³¹

One drawback to rater training was mentioned above: the reduction in determinant validity and expansion of interrater variance after lectures about avoiding halo error. Another consideration is that the impact of training effects on subsequent performance appraisals does not last long. The most commonly cited "washout" period is one week.³² The essential point is that rater errors appear to be well-developed habits resistant to lectures, practice, or warning.³³

Rating Formats. Trait checklists for grading complex personal characteristics, e.g., cooperativeness and independence, are looked upon unfavorably in management circles and by the courts. From an organizational standpoint, the main problem with trait measures is poor feedback: telling someone to "show more initiative" or to "be a better listener" does little good

without telling the person how to implement the advice in behavioral terms.³⁴ From a legal perspective, personnel decisions based on trait measures and affecting protected classes (e.g., minorities) are deemed "susceptible to partially and to the personal taste, whim, or fancy of the evaluator."³⁵

These problems are well known. What is less widely recognized is that trait measures and other ad hoc instruments have been shown to be no worse than, and most often superior to, behavioral scaling strategies. In one study of police job performance, three rating instruments—a trait checklist, a behaviorally anchored rating scale (BARS), and a simpler numerical rating scale—were compared with respect to several criteria. Although no instrument was clearly superior for all purposes, the results taken together suggest the following ranking from best to worst: the numerical rating scale, trait ratings, and the BARS format.³⁶ This ordering of the relative superiority of the instruments is contrary to theoretical expectations in a number of fields but is in general agreement with the reviews of the literature bearing on empirical testing of scaling strategies.³⁷

Trait measures continue to be the most widely used type of performance appraisal method.³⁸ By one estimate, 90 percent of today's organizations evaluate an employee's effectiveness in terms of personal qualities thought desirable by those in the organizational hierarchy.³⁹

MBO. Management by objectives (MBO) and performance appraisal are often confused. The key to MBO's success is goal setting tailored to the individual, while performance appraisal attempts to apply uniform standards to all employee behavior. Nevertheless, writers have commonly adopted two interconnected postulates that hold first, that MBO leads to higher performance; and second, that performance appraisal should focus on goal attainment.

Underlying the first postulate is the belief that participative goal setting is superior to assigning goals. The research on this issue is equivocal. It would appear that goal specificity and goal acceptance

can be attained as easily through assigned as through mutually set goals but that the establishment of higher goals is enhanced more by employee participation than by assigning higher goals or telling employees to do their best.⁴⁰ Although there is a heavy emphasis in the literature on mutual goal setting for the sake of professional employees, one review article found that such participation is "most effective among uneducated and minority-group members, perhaps because it gives them a sense of control over their fate."⁴¹

The second postulate, that performance raters should emphasize goal attainment, ignores the maxim of attribution theory: for more effective appraisals, raters should focus on behavior and not outcome. Research has shown that if employees are evaluated on goal attainment, "they are likely to set low goals or reject higher goals imposed by supervisors," but that if employees are evaluated on performance, "they will continue to set high goals whether the goals are attained or not."⁴²

Again, the role of participation is unclear. When one of the parties in mutual goal setting is "more equal" than the other (as is the norm), the question of management by whose objectives inevitably arises.⁴³ Although quantification is not possible, observers seem to agree that MBO includes a lot of spurious participation, in which the fiction is maintained that the subordinate is making a real input into work planning and goal setting.⁴⁴

Moreover, MBO is not for everyone. Job experience is necessary before participation makes much sense. For new employees, as well as troublesome ones and those who do not wish to participate, a more appropriate personnel style than MBO is what behavioralists call the *watcher*. This means getting individual employees to perform the way you want by reminding them that you are overseeing the situation.⁴⁵

Interview. During the appraisal interview, the supervisor and the subordinate are generally working at cross purposes. Supervisors have to make judgments affecting the subordinates' careers, morale, and relative share of rewards. Subordi-

nates need to verify their self-esteem and also obtain enough rewards. Such conflicts tend to generate ambivalent patterns of avoidance and defensiveness.⁴⁶

Mutual avoidance of the risks of negative feedback has given rise to a process called the "vanishing appraisal," in which both parties implicitly collude, in effect, to keep matters short and sweet.⁴⁷ Defensiveness springs from criticism of any sort.⁴⁸ To be effective at all, negative feedback should focus on specific behavioral changes to improve job performance; but there is no hard evidence that well-intentioned "constructive criticism" actually works.⁴⁹

Research in this area focuses on the relationship between appraisal interview characteristics and subsequent job performance. At least six major characteristics of an effective appraisal interview are well documented (although it is unclear whether they are all conceptually distinct). These are (1) the subordinate's prior preparation for the interview, (2) his or her opportunity to present ideas and feelings, (3) the degree of helpfulness of the supervisor, (4) the extent to which job problems are cleared up, (5) the extent to which goals are set, and (6) the absence of threats. Each characteristic has been found to be significantly related to improvement in later job performance, as well as to satisfaction with the appraisal process.⁵⁰

II. THE MINIMAL APPROACH

The nonprofit organization faces a special dilemma in the appraisal process not ordinarily found in the business or governmental sectors. The area of this dilemma, noted in a national survey of the quality of employment,⁵¹ is located between work characteristics and organizational climate. On the one hand, the survey found that "nonprofit employees, less fettered by centralization and control, have more autonomy in doing their jobs, report more variety and challenge, and find that their education is matched to their job demands." The cost of such job autonomy and employee influence is lack of job feedback, whether from the work itself or from supervisors and peers.

On the other hand, as the survey also showed, nonprofit organizations generally attempt to draw together their employees by democratic arrangements, so that participation "may create a greater sense of commitment and involvement for employees and serve to clarify jobs and soften role stresses." The survey concluded that efforts to import more businesslike performance appraisal systems into nonprofit organizations for the sake of job feedback could "centralize authority, limit flexibility, stifle innovation, and . . . alienate employees not only from their service but also from their ideals."⁵²

This argument parallels the general call in some quarters to make performance appraisals, as an ethical matter out of fairness, less ambitious and important.⁵³ Although no "minimal model" has been drafted, the basic position is that managers should only attempt to identify extremes of performance and leave salary administration to impersonal forces, such as the seniority system.

Apart from likely violation of legal ground rules, there are two fundamental weaknesses with the minimal approach. One is that the large area of indeterminacy in appraisals is an important fiduciary component of the supervisor's role, not an excuse to skip or minimize the appraisal process. On the whole, employees want valid job feedback. They may not be looking for criticism, but they do not want to spend their professional lives uninformed about the character and results of their work. The other weakness of the minimal approach is that it downplays the need for goal setting and job feedback, and apparently ignores the "Pygmalion effect," which has shown again and again that people perform best when they have high supervisor expectations, encouragement, and visual attention.

III. HEURISTIC KNOWLEDGE

The heuristic side of performance appraisal consists of strategies and rules of thumb for deciding which among several alternative courses of action promises to be the most expedient one. In the mathematical and physical sciences, some heu-

ristic rules can leap across logical gaps in problem solving. However, in immature or ineffective fields such as social sciences, heuristic knowledge consists of intuitive generalities, or aphorisms, which have neither the objectivity, nor the permanence of an impersonal law.

The heuristic potential of any field depends on whether the important problems in that field "can be effectively solved to yield some sort of genuine knowledge."⁵⁴ Study of the appraisal process is a venture into tremendously complicated problems of organizational behavior. *Behavior* includes not only overt acts but also subjective behavior, such as attitudes, beliefs, motivations, and desires. No significant empirical or theoretical progress has been made in this field for a long time. A major review of the 1970s literature had little positive to report on performance appraisal: "Research convincingly showed that instrumentation [i.e. a behavioral rating scale] has minimal impacts on psychometric errors, but that conclusion was pretty well established before."⁵⁵ A more current assessment of research work in organizational psychology concluded that "the theory/practice gap will close by a decrease in the relevance of theory, not by an increase in the quality of the practice."⁵⁶

About the only strong consensus with empirical support in the heuristic literature on performance appraisal deals with the fundamental importance of setting goals and giving feedback. Together, these are not only more effective than alternative methods, such as participatory decision making or job enrichment, but may also be the major mechanisms by which these other incentives affect motivation. Even money is a more effective motivator when bonuses are made contingent on the attainment of specific objectives.⁵⁷

But on issues of technique having a direct bearing on how to set goals and give feedback—for example, the relative merit of assigned versus participatory goals, or of intrinsic versus extrinsic rewards, or of weekly versus annual appraisal sessions—theorists plainly disagree.

The potential development of heuristic rules about performance appraisal is undetermined not only by the contemporary scarcity of empirical generalizations, but also by two chronic limitations: the valuing aspects of appraisal, and common sense.

In moving along a continuum from behavior-based, *mechanistic* tasks to those that are judgment based and, hence, *organic* in structure, the valuing aspect of appraisal becomes increasingly salient. Even for moderately uncertain tasks, "articulation of a mutually agreeable standard of measurement is normally not possible."⁵⁸ A series of studies on this shift from measurement to valuing as appraisal uncertainty increases demonstrated that value is likely to be assigned on the basis of "personal factors of relevance," and concluded that personal judgments should be regarded as "critical and legitimate concerns" in those situations "where controversy and interest are joined."⁵⁹

As a general surmise, common sense is not an infallible guide to performance appraisal. Herbert White⁶⁰ and Richard De Gennaro,⁶¹ among other writers, rightly argue for the value of experience, the proper attitude, and intelligence in library management; but Peter Drucker singles out two areas in which such tacit knowledge is not enough. "One is personnel decisions, the other is marketing."⁶² Drucker's statement cannot be "proved," but there is supportive research on the nature of tacit knowledge (discussed below); and an old but unchallenged review of the literature concluded that the inability to judge the work traits and aptitudes of others is itself a personality trait.⁶³

Behavioral study of performance appraisal is not simply stalled, nor its heuristic knowledge merely ineffective. Given both a lack of empirical generalizations and a propensity for subjective evaluations and personal feelings, the prevailing heuristic literature resembles a *folk-science*. This is "a body of accepted knowledge whose function is not to provide the basis for further advance, but to offer comfort and reassurance to some body of believers."⁶⁴

IV. IN-HOUSE EXPERIMENTATION

Broadly defined, the in-house experiment is a nonrandom, quasi-empirical plan "to generate knowledge about human options that go beyond the status quo."⁶⁵ Its nonrandom character derives from the self-selected population (of participants). Its quasi-empirical character involves a shift from positivist logic (normal science) to tacit knowledge, which is partial and subjective. *To generate knowledge* is not simply to make an inquiry; *knowledge* is the capacity to detect and deal with error. Finally, *human options* can apply to various aspects of organizational behavior: work motivation, staff productivity, job enrichment, and participatory management, as well as performance appraisal.

Any stricter definition of the in-house experiment would be arbitrary. The philosophical pluralism now current in the field of organizational psychology is one of its chief sources of strength,⁶⁶ but it is also a broad and general feature; and the paucity of published reports on this type of experiment makes it difficult to find models or paradigms.⁶⁷ The library science literature, for example, is full of narratives about performance appraisal but only one article stands out as a true experiment, in which a definite plan leads to action that detects and attempts to correct a mismatch between intentions and organizational outcomes.⁶⁸

Reports on in-house experimentation have been made, for the most part, by a small band of theorist-practitioners who have tied adult learning strategies to the study of organizational culture.⁶⁹ Their findings suggest two general reasons why an in-house experiment is not an ordinary plan of action: "the inability of individuals to learn the skills to produce conditions of trust, candidness, and risk-taking under online conditions, and the lack of organizational support for such changes."⁷⁰ With reference to the first reason, a fairly sophisticated study of the types of "master plans" people hold about managing their actions determined that "individuals who are young or old, male or female,

white or black, rich or poor, powerful or powerless, may hold different *espoused* theories, they may *act* differently, but their interpersonal *theories-in-use* do not seem to vary." A model of this more or less universal behavioral pattern has four objectives: "(1) control the purpose of the situation, (2) minimize losing and maximize winning, (3) suppress negative feelings, and (4) emphasize rationality."⁷¹

Additionally, research work on leadership points up a marked tendency for organizational administrations to avoid in-house experiments. Supervisors at every level tend to have a basic lack of confidence in others and do not believe that self-management traits—initiative, judgment, responsibility—are widely distributed among subordinates.⁷² And a clearly large number of supervisors hold to a lump-of-control theory. This type of theory-in-use assumes that there is a virtually fixed amount of authority in any organization, so that whatever the supervisor surrenders is a political and even a personal loss.⁷³

All this suggests behavioral reasons why in-house experiments to alter the status quo have seldom been put into action. On a theoretical plane, it seems unlikely that such experiments could evolve into strategic sources for model building because their methods and criteria consist as much of independent variables as of organizational behavior. Even MBO, in practice, has scores of variations. Accordingly, the confidence level of an in-house experiment may decay quickly when findings are removed from the original setting.⁷⁴

An associated theoretical issue centers on the intrinsically mental character of in-house experiments (and of all social phenomena).⁷⁵ Several labels for personal knowledge about organizational behavior in a particular setting have appeared (e.g., indwelling understanding,⁷⁶ armchair theorizing),⁷⁷ but most writers have adopted Michael Polanyi's 1972 term, *tacit knowledge*.⁷⁸

Everyone uses tacit knowledge to make sense of the interplay of intentions, interactions, and context in daily life. In an in-house experiment, the role of tacit

knowledge—experience, intuition, intelligence, guesswork—would be preeminent in designing the approach and in selecting and interpreting the results:

The research ought to be skeptical of positivist science when the unit of analysis is, like the researcher, a self-reflecting subject, when relationships between subjects (actors) are influenced by definitions of the situation, or when the reason for undertaking the research is to solve a problem which the actors have helped to define.⁷⁹

Such action research makes theorists nervous because reliance on a participant's personal assessment implies a profundity that can be self-aggrandizing, defensive, or misinformed. To temper and round out subjective assessments, writers uniformly recommend the panel or team design for evaluation of nonrandom in-house experiments.⁸⁰

CONCLUSIONS

Performance appraisal provides a modicum of rationality and legality to key personnel actions, but there is no neat, objective way of going about it because of common, intractable sources of rater error. Of the four approaches to performance appraisal under review—behavioralism, minimal model, heuristic knowledge, and in-house experimentation—only the last one has the capacity to detect and deal with error. All the other approaches are of relatively limited utility.

The essential difficulty with *behavioralism* is that our state of knowledge about the determinants of appraisal effectiveness lags far behind organizational needs and legal requirements. Behaviorally based methods, when evaluated for validity, relevance, discriminability, and freedom from bias, remain inadequate for the purposes that appraisals are intended to serve. This all-too-apparent deficiency in personnel administration has come under increasing legal scrutiny,⁸¹ but there is little documentation of the particular laws governing appraisals in small nonprofit organizations.

No dramatic breakthroughs in the theory or practice of performance appraisal are in sight. Practically all of the concepts and methods now in use have been

around for decades, and the types of method still to be tested do not appear promising.⁸² Thus, the search for an ultimate rating format was plainly exhausted in the 1970s when researchers quit hedging about the efficacy of behavioral scaling strategies. Rater training programs may be on the decline, but misapplications of MBO as a grand substitute for an appraisal system will likely endure. At the theory-building level, a workable model of the appraisal process would encompass such a vast array of human, social, and organizational variables that, if feasible, would require decades to develop.⁸³

The antithesis of behavioralism is paternalism, involving a *minimal model* with two chief premises: benign neglect of performance appraisal; and "unconditional rewards in the sense that the amount of reward that any individual receives is not dependent in any clear-cut way on how he behaves within the organization, but rather on the fact that he is a member of that organization."⁸⁴ Although this model has been eclipsed in the management literature since the mid-1950s by others (e.g., MBO, participatory decision making, behavioralism), it is still salient in the nonprofit sector and especially in the library field.

Relaxation of performance appraisal and resort to across-the-board pay increases may be entirely reasonable in some organizations at some times, such as during periods of budgetary restraint. But, over the long haul, the premises of paternalism tend to break down. First, where performance appraisals are not conducted, "implicit evaluations will be assumed, and what managers say and do, even casually, will be interpreted and weighted as organizational evaluations."⁸⁵ Second, an unconditional rewards system based on membership and egalitarianism may well be dissatisfying not only to high achievers but also to many average workers, given the high inflation of self-ratings relative to others' performance.

The aim of *heuristic literature* is to disseminate usable knowledge in a particular field. In the social and behavioral sciences, the logic of positivism—that human interaction and society can be understood in

terms of impersonal laws—has led to research reports that are quite narrow in scope and technical in application. One argument of an emergent postbehavioral movement is that scientific certainty of knowledge, to a degree, is the enemy of its usefulness:

The effort to establish certainty almost always leads to large numbers of studies being done on a single small topic and to more and more careful specification of the phenomenon. Once the phenomenon has been subjected to all the tests of certainty, it often winds up so complex that it is no longer useful to the practitioner.⁸⁶

Hence, the important function of heuristic literature is to make scientific knowledge understandable and accessible to the public. This literature holds few empirically grounded generalizations about performance appraisal, but it certainly provides readier access to useful knowledge than could be obtained from a survey of the primary research in behavioral science journals.

Analysis of the several main approaches to performance appraisal along the foregoing lines leads to the inevitable conclusion that performance appraisal is not an effective field. However, this does not require us to assume that nothing can be done, through deliberate plan, to improve the appraisal process. It seems reasonable to expect that real improvements can be achieved through a good measure of in-house experimentation and some basic changes in the way we deal with unspoken assumptions about our ability to judge the work of others.

Unlike rater errors in performance appraisal, the problems of *in-house experimentation* are not intractable. The sine qua non is "double-loop learning,"⁸⁷ which occurs when, in order to deal with error, behavioral values and tacit knowledge are questioned and altered. (Single-loop learning, by contrast, involves the detection of only routine errors that do not upset the underlying values of the theory-in-use).

The elusive goal of in-house experimentation has been called reflexive power,⁸⁸ or the enlightenment effect, whereby knowledge of behavioral laws can modify or even nullify their impact on those who

share that knowledge.⁸⁹ However, discussions in the literature about the enlightenment effect are necessarily abstract because such tacit knowledge is "a-critical"⁹⁰—it cannot be articulated expressly. Thus, it does not pass easily as public (or heuristic) knowledge:

Even if we begin to have an intuitive understanding of an organization's culture, we may find it extraordinarily difficult to write down that understanding in such a way that the essence of the culture can be communicated to someone else.

... But when we see the essence of a culture, the paradigm by which people operate, we are struck by how powerful our insight into that organization now is, and we can see instantly why certain things work the way they do, why certain proposals never get bought, why change is so difficult, why certain people leave, and so on.

... It is the search for and the occasional finding of this central insight that makes it all worthwhile . . . even if, in the end, we can share it only with colleagues.⁹¹

In brief summary, perhaps an appropriate skepticism of the approaches to performance appraisal other than in-house experimentation is the comedian's dictum, "It isn't ignorance that does the most damage, it's knowing for sure so much that isn't so." These approaches involve only "single-loop learning" and do not have the capacity to detect and deal with basic error. In-house experimentation, by contrast, is a way of correcting overlearning without overintellectualizing the problem.

In using nonexperimental approaches we judge the work and potential of others with a set of behavioral values and tacit theories-in-use. These may or may not be accurate or corrigible, but most organizations have never thought through what they regard to be desirable about personal traits and organizational behavior, what relative importance they give to the various values they hold, or how these values can be incorporated into the current appraisal system. We would be better off if we made a serious endeavor to assess and experiment with our theories-in-use, starting with more doubts about what we know "for sure."

REFERENCES

1. Thomas DeCotiis, "An Analysis of the External Validity and Applied Relevance of Three Rating Formats," *Organizational Behavior and Human Performance* 19:265 (Aug. 1977); Jeffrey Kane and Edward Lawler III, "Performance Appraisal: Its Assessment and Determinants," in *Research in Organizational Behavior: An Annual Series of Analytical Essays and Critical Reviews*, ed. by Barry Staw, V.1 (Greenwich, Conn.: JAI Press, 1979), p.456-58.
2. Douglas McGregor, "An Uneasy Look at Performance Appraisal," *Harvard Business Review* 53:89-94 (May-June 1957).
3. Dimitry Berkner, "Library Staff Development through Performance Appraisal," *College and Research Libraries* 40:335-44 (July 1979); G. Edward Evans, *Management Techniques for Librarians*, 2d ed. (New York: Academic, 1983) p.211-43; David Kaser and Ruth Jackson, "A Century of Personnel Concerns in Libraries," in *A Century of Service: Librarianship in the United States and Canada*, ed. by Sidney L. Jackson, Eleanor B. Herling, and E. J. Josey (Chicago: American Library Assn., 1976), p.129-45; H. R. Kroll, "Beyond Evaluation: Performance Appraisal as a Planning and Motivational Tool in Libraries," *Journal of Academic Librarianship* 9:27-32 (Mar. 1983); Murray S. Martin, *Issues in Personnel Management in Academic Libraries* (Greenwich, Conn.: JAI Press, 1981), p.153-57; Maxine Reneker, "Performance Appraisal in Libraries: Purpose and Techniques," in *Personnel Administration in Libraries*, ed. by Sheila Creth and Frederick Duda (New York: Neal-Schuman, 1981) p.227-89; Herbert S. White, *Library Personnel Management* (White Plains, N.Y.: Knowledge Industry Pub., 1985), p.123-38.
4. Philip Mirvis and Edward Hackett, "Work and Work Force Characteristics in the Nonprofit Sector," *Monthly Labor Review* 15:3-12 (Apr. 1983).
5. Peter F. Drucker, *The Age of Discontinuity: Guidelines to Our Changing Society* (New York: Harper, 1969).
6. Kim Cameron, "Critical Questions in Assessing Organizational Effectiveness," *Organizational Dynamics* 9:66-80 (Autumn 1980).
7. Randall Schuler, *Personnel and Human Resources Management*, rev. ed. (New York: West Pub., 1984), p.271-79.
8. Elaine Gruenfeld, *Performance Appraisal: Promise and Peril* (Ithaca, N.Y.: New York State School of Industrial and Labor Relations, Cornell Univ., 1981), p.18-20; Kane, "Performance Appraisal," p.452-54.
9. W. Kirchner and D. Reisberg, "Differences between Better and Less Effective Supervisors in Appraisals of Subordinates," *Personnel Psychology* 15:295-302 (Autumn 1962).
10. Gruenfeld, *Performance Appraisal*, p.10-13.
11. Wayne Cascio and John Bernardin, "Implications of Performance Appraisal Litigation for Personnel Decisions," *Personnel Psychology* 34:215 (Summer 1981).
12. Walter Borman, "Consistency of Rating Accuracy and Rating Errors in the Judgment of Human Performance," *Organizational Behavior and Human Performance* 20:238-52 (Dec. 1977); Robert Holzbach, "Rater Bias in Performance Ratings; Superior, Self-, and Peer Ratings," *Journal of Applied Psychology* 63:579-88 (Nov. 1978); William Scott, "The Influence of Variations in Performance Profiles on the Performance Evaluation Process: An Examination of the Validity of the Criterion," *Organizational Behavior and Human Performance* 14:360-70 (Dec. 1975).
13. Walter Borman, "Effects of Instructions to Avoid Halo Error on Reliability and Validity of Performance Evaluation Ratings," *Journal of Applied Psychology* 60:556-60 (May 1975); Holzbach, "Rater Bias," p.581.
14. Borman, "Effects of Instructions," p.557.
15. Borman, "Effects of Instructions," p.560.
16. Walter Borman, "Exploring Upper Limits of Reliability and Validity in Job Performance Ratings," *Journal of Applied Psychology* 63:142 (May 1978).
17. Gary Latham, Larry Cummings, and Terrence Mitchell, "Behavioral Strategies to Improve Performance," *Organizational Dynamics* 9:5-23 (Winter 1981). Terrence Mitchell and Robert Wood, "Supervisors' Responses to Subordinate Poor Performance: A Test of an Attributional Model," *Organizational Behavior and Human Performance* 25:123-38 (Feb. 1980).
18. Latham, "Behavioral Strategies," p.14.
19. Mitchell, "Supervisors' Responses," p.138.
20. Jess Martin, "Staff Evaluation of Supervisors," *Special Libraries* 70:26-29 (Jan. 1979).
21. Holzbach, "Rater Bias," p.579-88. Benson Rosen and Thomas Jerdee, "Sex Stereotyping in the

- Executive Suite," *Harvard Business Review* 52:43-56 (Mar.-Apr. 1974).
22. Marsha Jacobson and Walter Koch, "Women as Leaders: Performance Evaluation as a Function of Method of Leader Selection," *Organizational Behavior and Human Performance* 20:149-57 (June 1977).
 23. Holzbach, "Rater Bias," p.580.
 24. Arie Lewin and Abram Zwyany, "Peer Nominations: A Model, Literature Critique and a Paradigm for Research," *Personnel Psychology* 29:423-47 (Aug. 1976).
 25. Lewin, "Peer Nominations," p.431-35.
 26. Schuler, *Personnel*, p.272-73.
 27. Herbert Heneman and Donald Schwab, "Assessing Outcomes," in *Perspectives on Personnel/Human Resource Management*, rev. ed. by Herbert Heneman and Donald Schwab (Homewood, Ill.: Irwin, 1982) p.68; Holzbach, "Rater Bias," p.580.
 28. Meyer, "Self-Appraisal of Job Performance," *Personnel Psychology* 33:293 (Summer 1980).
 29. Borman, "Effects of Instructions," p.556.
 30. Cascio, "Implications," p.223.
 31. Nathan Winstanley, "Performance Appraisal: Another Pollution Problem?" *Conference Board Record* 9:59-63 (Sept. 1972).
 32. Dennis Warmke and Robert Billings, "Comparisons of Training Methods for Improving the Psychometric Quality of Experimental and Administrative Performance Ratings," *Journal of Applied Psychology* 64:129 (Feb. 1979).
 33. Latham, "Behavioral Strategies," p.19.
 34. Latham, "Behavioral Strategies," p.7.
 35. *Wade v. Mississippi Cooperative Extension Service*, 528 F.2d 416 (7th Cir. 1978).
 36. DeCotiis, "An Analysis."
 37. Wayne Cascio and Enzo Valenzi, "Behaviorally Anchored Rating Scales: Effects of Education and Job Experience of Raters and Ratees," *Journal of Applied Psychology* 62:278-82 (Aug. 1977); DeCotiis, "An Analysis"; Gruenfeld, *Performance Appraisal*, p.13-18; Donald Schwab, Herbert Heneman III, and Thomas DeCotiis, "Behaviorally Anchored Rating Scales: A Review of the Literature," *Personnel Psychology* 28:549-62 (Winter 1975).
 38. Barry Friedman and Edwin Cornelius III, "Effect of Rater Participation in Scale Construction on the Psychometric Characteristics of Two Rating Scale Formats," *Journal of Applied Psychology* 61:210-16 (May 1976).
 39. Latham, "Behavioral Strategies," p.7.
 40. Gary Latham, Terence Mitchell, and Dennis Dossett, "Importance of Participative Goal Setting and Anticipated Rewards on Goal Difficulty and Job Performance," *Journal of Applied Psychology* 63:163-71 (May 1978).
 41. Gary Latham and Edwin Locke, "Goal Setting—A Motivational Tool That Works," in *Psychological Foundations of Organizational Behavior*, rev. ed. by Barry Staw (Glenview, Ill.: Scott, Foresman, 1983), p.252.
 42. Latham, "Importance of Participative Goal Setting," p.169-70; Latham, "Behavioral Strategies," p.20.
 43. Frederick Herzberg, "One More Time: How Do You Motivate Employees?," *Harvard Business Review* 46:61 (Jan.-Feb. 1968); Craig Pinder, "Concerning the Application of Human Motivation Theories in Organizational Settings," *Academy of Management Review* 2:384-97 (July 1977).
 44. Michael Beer, "Performance Appraisal: Dilemmas and Possibilities," *Organizational Dynamics* 9:28 (Winter 1981); Schuler, *Personnel*, p.270.
 45. Ron Kolgraf, "Psychology and Selling," *Industrial Distribution*, p.39 (Oct. 1978).
 46. Beer, "Performance Appraisal," p.28; Schuler, *Personnel*, p.270.
 47. Beer, "Performance Appraisal," p.27.
 48. Ronald Burke, William Weitzel, and Tamara Weir, "Characteristics of Effective Employee Performance Review and Development Interviews: Replication and Extension," *Personnel Psychology* 31:903-11 (Winter 1978).
 49. Schuler, *Personnel*, p.270.
 50. Gruenfeld, *Performance Appraisal*, p.22-27; Burke, "Characteristics," p.915-916.
 51. Mirvis, "Work and Work Force Characteristics."
 52. Mirvis, "Work and Work Force Characteristics," p.11.
 53. Gruenfeld, *Performance Appraisal*, p.23; Schuler, *Personnel*, p.290; N. Winstanley, "Legal and Ethical Issues in Performance Appraisals," *Harvard Business Review* 58:186-88 (Nov.-Dec. 1980).
 54. Jerome R. Ravetz, *Scientific Knowledge and Its Social Problems* (Oxford, Eng.: Clarendon, 1971), p.377.

55. Lee Dyer and Donald P. Schwab, "Personnel/Human Resource Management Research," in *Industrial Relations Research in the 1970s: Review and Appraisal*, ed. by Thomas A. Kochan, Daniel J. B. Mitchell, and Lee Dyer (Madison, Wis.: Industrial Relations Research, 1982), p.194.
56. J. Richard Hackman, "Doing Research That Makes a Difference," in *Doing Research That Is Useful for Theory and Practice*, ed. by Edward E. Lawler III and others (Washington: Jossey-Bass, 1985), p.150.
57. Latham, "Goal Setting," p.45-46.
58. Michael Keeley, "A Contingency Framework for Performance Evaluation," *Academy of Management Review* 3:398-99 (July 1978).
59. Keeley, "A Contingency Framework," p.403.
60. White, *Library Personnel Management*.
61. Richard De Gennaro, "Theory vs. Practice in Library Management," *Library Journal* 108:1318-21 (July 1983).
62. Peter F. Drucker, "Managing the Public Service Institution," *College & Research Libraries* 37:4-14 (Jan. 1976).
63. Ronald Taft, "The Ability to Judge People," *Psychological Bulletin* 52:1-23 (Jan. 1955).
64. Ravetz, *Scientific Knowledge and Its Social Problems*, p.366.
65. Chris Argyris, "Theories of Action That Inhibit Individual Learning," *American Psychologist* 31:638 (Sept. 1976).
66. Nigel Nicholson and Toby Wall, "Introduction: Themes in the Theory and Practice of Organizational Psychology," in *The Theory and Practice of Organizational Psychology: A Collection of Original Essays*, ed. by Nigel Nicholson and Toby D. Wall (New York: Academic, 1982), p.12.
67. Edgar H. Schein, *Organizational Culture and Leadership* (Washington: Jossey-Bass, 1985), p.147.
68. Martin, "Staff Evaluation of Supervisors."
69. Chris Argyris and Donald Schon, *Organizational Learning: A Theory of Action Perspective* (Reading, Mass.: Addison-Wesley, 1978); Chris Argyris, *Reasoning, Learning, and Action: Individual and Organization* (Washington: Jossey-Bass, 1982); Donald A. Schon, *The Reflective Practitioner* (New York: Basic Books, 1982); David A. Kolb, *Experimental Learning: Experience as the Source of Learning and Development* (Englewood Cliffs, N.J.: Prentice-Hall, 1984).
70. Chris Argyris, "Research as Action: Usable Knowledge for Understanding and Changing the Status Quo," in *The Theory and Practice of Organizational Psychology*, p.206.
71. Argyris, "Research as Action," p.200. Emphasis added.
72. Raymond E. Miles, "Human Relations or Human Resources?" in *Organizational Psychology: A Book of Readings*, ed. by David A. Kolb, Erwin M. Rubin, and James M. McIntyre (Englewood Cliffs, N.J.: Prentice-Hall, 1971), p.238.
73. Miles, "Human Relations or Human Resources?," p.237-238.
74. Frank Landy, "Models of Management: Assumptions of Theorists," in *The Theory and Practice of Organizational Psychology*, p.103-21; Jerald Hage, "Theory-building," in *The Theory and Practice of Organizational Psychology*, p.123-47.
75. John Searle, *Minds, Brains and Science* (Cambridge, Mass.: Harvard Univ. Pr., 1984), p.84.
76. Nicholson, "Introduction," p.15.
77. Nicholson, "Introduction," p.12.
78. Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy* (Chicago: Univ. of Chicago Pr., 1972), p.373.
79. Gerald I. Susman and Rodger D. Evered, "An Assessment of the Scientific Merits of Action Research," *Administration Science Quarterly* 23:600 (Dec. 1978).
80. Hackman, "Doing Research That Makes a Difference," p.154; Donald A. Schon, "Education for Reflection-in-Action: An Alternative to the Positivist Epistemology of Practice," in *The Optimum Utilization of Knowledge: Making Knowledge Serve Human Betterment*, ed. by Kenneth E. Boulding and Lawrence Senesh (Boulder, Colo.: Westview Pr., 1983), p.314-22; William R. Ferrell, "Combining Individual Judgments," in *Behavioral Decision Making*, ed. by George Wright (New York: Plenum Pr., 1985), p.111-45.
81. Robert Lazer, "The 'Discrimination' Danger in Performance Appraisal," *Conference Board Record* 13:160 (Mar. 1976).
82. Kane, "Performance Appraisal," p.442.
83. Kane, "Performance Appraisal," p.426.
84. Edward E. Lawler III, *Motivation in Work Organizations* (Monterey, Calif.: Brooks/Cole, 1973), p.199.
85. Albert Shapero, *Managing Professional People: Understanding Creative Performance* (New York: Free Pr., 1985), p.90.

86. Edward E. Lawler III, "Challenging Traditional Research Assumptions," in *Doing Research That Is Useful for Theory and Practice*, p.12.
87. Argyris, "Research as Action," p.201.
88. Nicholson, "Introduction," p.3.
89. Kenneth J. Gergen, "Social Psychology as History," *Journal of Personality and Social Psychology* 26:309-20 (May 1973).
90. Polanyi, *Personal Knowledge*, p.264.
91. Schein, *Organizational Culture and Leadership*, p.147.