Performance-Appraisal Accuracy: 
An Illusive or Sometimes 
Misguided Goal?

Daniel R. Ilgen 
Michigan State University

BACKGROUND

Approaches to Performance Appraisal

There are few statements more central to the practice of industrial and organizational psychology than the adages to "know thy job" and "know thy people." In the latter case, the rule is often institutionalized in the form of formal performance-appraisal systems. The implicit goal of all such systems is to provide accurate appraisals of employee performance based on the judgments of one or more other persons who are also organizational employees.

Unfortunately, rarely if ever is anyone satisfied with such ratings. They never seem to meet anyone's expectations. Those being appraised are satisfied only if the appraisals match their often inflated beliefs about their own performance. Those who complete the evaluations find them time consuming, bothersome, and disquieting when they must provide performance feedback. Managers and others who look at the appraisal data of many employees often find them less than credible as they try to reconcile the apparent discrepancy between the glowing appraisal reports on file and their own personal beliefs about the level of performance in the work force.

Dauntless, we industrial and organizational psychologists see the problems not as debilitating limitations but as challenges to be met. Given our psychometric training and our knowledge, capabilities, and demonstrated success developing systems for assessing jobs and individual differences, our initial response
to such problems was to assume that the limitations were rooted in rating errors and other manifestations of measurement problems. This led to extensive work on the definition of criteria of job performance and the development of measures of these criteria. Our early efforts were focused on three interrelated sets of issues. The first was the development and evaluation of scales for rating performance that were psychometrically sound. The second was training evaluators to use the performance-rating instruments effectively. Such training focused on insuring familiarity with the scales themselves, standardizing their use across raters in the sample of raters, creating an awareness of the types of errors to which raters often fall victim, and then providing ways to attempt to avoid these errors. Finally, at the system level, policies and practices complete with manuals, schedules of when to appraise, and centralized procedures to keep track of the appraisal process were developed.

By almost any standard, our efforts were successful. We identified performance-appraisal scales that are better than others under particular conditions (Bernardin & Beatty, 1984); we learned how to train raters to reduce rating errors and improve their ability to use the performance-appraisal system (Wexley & Latham, 1981); and we established guidelines for the development and implementation of appraisal systems (see, for example, Bernardin & Beatty, 1984). In spite of these gains, the dissatisfaction with performance appraisals remained.

Landy and Farr, in their watershed 1980 Psychological Bulletin article, reviewed the literature on performance appraisal up to that time and responded to the criticisms by suggesting that attention shift from a concern for the rating scale, training, and appraisal practices related to using particular scales to that of understanding the judgment processes of raters. These authors suggested that work in the late 1970s combining cognitive processing with person perception provided provocative ideas about the ways in which persons who rate the performance of others may observe, store in memory, and recall social cues that form the bases for evaluations of others. At about the same time, Jack Feldman (1981) was even more specific in linking social cognition with performance appraisal. A flurry of research activity developed pursuing the understanding of the cognitive processes of performance appraisers. This research activity dominated performance appraisal research in the 1980s.

Almost without exception, the primary criterion of the recent research has been some index of the extent to which raters accurately report the behaviors of ratees or accurately evaluate ratee performance. That is to say, in earlier research, it was assumed that the goal of the appraisal process was to provide accurate observations and evaluations of employees, but research rarely addressed rater accuracy directly. If accuracy was addressed, the tendency was to assume that rating errors were indicators of accuracy or that intrarater agreement represented accuracy. The work in the 1980s questioned both rating errors and intrarater agreement as measures of rater accuracy, preferring instead
the establishment of known standards of performance in which to compare raters' appraisals. To create conditions where the standards of performance were known, it was usually necessary to conduct research under far more controlled conditions than had been done for most performance-appraisal research prior to 1980. Much of the later research was done in the laboratory.

In a recent review and evaluation of the performance-appraisal literature of the 1980s directed at understanding factors affecting raters' ability to provide accurate ratings, Ilgen, Barnes-Farrell, and McKellin (in press) concluded that the research had contributed to a better understanding of four general issues related to the application of performance appraisals in organizations. These four related to: (a) the need to improve the raters' ability to obtain unbiased samples of ratee behaviors prior to making appraisal judgments, (b) the correction of misguided advice about the nature of rating scales and the ability of raters to observe events without simultaneously evaluating those events, (c) the importance of the frames-of-references of raters (called "schemas" in the jargon of social cognition), and (d) the need to be extremely cautious when using archival performance-appraisal ratings due to the effects of the purposes of ratings and other setting-specific variables on ratings that appear in personnel files.

Reactions to Performance Appraisals

Ironically, in spite of a number of major gains in the last 15 to 25 years in performance-appraisal technology, dissatisfaction with performance appraisals is no less today than it was in the past. If anything, criticisms have been more prevalent. For example, Longenecker, Sims, and Gioia (1987) reported the results of their lengthy interviews regarding performance appraisals with 60 executives representing 11 functional areas in 7 large corporations. The authors concluded that the political realities of corporate life supersede goals of accuracy and honesty when managers are asked to complete performance appraisals. Mohrman and Lawler (1983), commenting on the link between appraisals and pay, stressed that those who think that tying salary decisions to performance appraisals will decrease the probability of bad salary decisions soon find that both their salary structures and their performance evaluations are based on bad data. Banks and Murphy (1985) criticized much of the performance-appraisal research of the 1980s for paying too little attention to the practitioner concerns of (a) managerial commitment to the appraisal process, (b) the nature of communications between supervisors who serve as raters and their subordinates, (c) supervisors' performance feedback skills, and (d) their ability to clarify objectives for others. Finally, I was struck by the invariance of the complaints over time when I attended a panel discussion at the annual conference of the Society of Industrial and Organizational Psychology held in Boston in April of 1989. The panel discussion was entitled, "Real World Performance Appraisal: What do Scientists have to Offer?" Of the four panelists, two were from academia
and two were employed in corporations where they developed or were involved with the operation of the performance-appraisal systems. The concerns that the organizational representatives raised about the lack of help that they saw coming from research on performance appraisal could just as easily have been raised in 1949 as in 1989, in spite of the fact that performance-appraisal scales, rater training, and operational procedures are considerably better today than at that time.

The cursory historical review just outlined presents a curious phenomenon. Research and development regarding employee performance-appraisal systems document real improvements over the years, and yet, in spite of the improvements, the perceptions of those who use such systems have not changed; the same old problems that plagued the systems from their inception exist today with little change in magnitude. Why?

Three explanations come to mind. The first of these is that the advancements that have been made in performance-appraisal systems by those who do research on the systems have not been incorporated into systems in use. Without a doubt, there are a number of systems in operation that have serious technical flaws. It is not hard to find systems with graphic rating scales having no anchors other than the endpoints, and these scales are being used for general performance outcomes such as “overall job performance” or for a list of ill-defined traits. Even in cases where the systems have been very carefully defined and introduced into the organization, often they are poorly maintained so that, over time, new raters are required to use the systems with inadequate training. Unfortunately, complaints about the systems are not limited to settings where the systems have not applied state-of-the-art procedures. Under the best of conditions, the concerns that were raised earlier prevail.

A second explanation is that with improved techniques also come rising expectations. Dissatisfaction could exist in the face of improved practices if the persons using the systems had expectations that remained some constant level above what the system could deliver, regardless of the fact that the system might deliver more at time \( t + 1 \) than at time \( t \). Although this explanation cannot be ruled out, it also seems unlikely. First of all, those who have introduced improvements in rating scales, rater training, and other components of performance-appraisal systems, in general, have tended to show a prudent respect for the limits of their techniques. Furthermore, the content of the issues addressed by the technologically focused improvements are often independent of the issues with which complaints are raised. For example, improvements in rater training may focus on reducing rating errors, but the concerns of the users may be with the political factors that dictate the use of the rating scale.

A third possibility is that those who are most interested in designing research on performance-appraisal systems may have looked at too narrow a domain for addressing the problems of these systems. For the most part, the researchers have been industrial and organizational psychologists or human resource manage-
ment persons oriented toward microindividual-level views of the appraisal process. Given this orientation, it is easy to understand why such persons would look to the nature of measurement instruments, scaling, rating errors, and training to attempt to "fix" the system. Likewise, when the limits of these approaches were addressed and alternatives suggested, it is understandable that researchers, drawn from this pool, look to other psychological literatures, in this case social cognition, to move from the limited effectiveness plateau. I suggest that it is time to redefine the problem from a different perspective and to explore whether such a redefinition will lead to possible changes in approaches to performance-appraisal systems that address some of the concerns that have remained with us over the years.

A Functional Perspective on Performance Appraisal

A Description of Functions. Rather than search our psychological roots for explanations of performance appraisals, let us ask a more basic question. Why establish a performance-appraisal system in the first place? Table 15.1 provides one way to answer this question. The answer is based on the assumption that performance-appraisal systems are designed to meet the needs of particular constituencies, and that there are certain functions served by performance appraisals for these constituencies.

It is suggested in Table 15.1 that the constituencies served by a formal performance appraisal can be clustered into two mutually exclusive categories. The first represents the organization's goals. It is labeled organizations. Next, there are the persons whose performance is appraised. This set is labeled individual appraisees. The functions that the appraisal system serves the organization are again clustered into two sets. The first of these deals with the need to encourage

<table>
<thead>
<tr>
<th>Constituencies</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Organizations</td>
<td>A. Control through Human Resource Management</td>
</tr>
<tr>
<td></td>
<td>1. Communicate Performance Standards</td>
</tr>
<tr>
<td></td>
<td>2. Administer Rewards</td>
</tr>
<tr>
<td></td>
<td>3. Discipline &amp; Dismissals</td>
</tr>
<tr>
<td></td>
<td>B. Data for System Level Decisions</td>
</tr>
<tr>
<td></td>
<td>1. Criteria for:</td>
</tr>
<tr>
<td></td>
<td>Selection</td>
</tr>
<tr>
<td></td>
<td>Training Evaluation</td>
</tr>
<tr>
<td></td>
<td>2. Needs Analyses</td>
</tr>
<tr>
<td>II. Individual Appraisee</td>
<td>A. Learn</td>
</tr>
<tr>
<td></td>
<td>B. Obtain Rewards</td>
</tr>
<tr>
<td></td>
<td>C. Avoid Negative Sanctions</td>
</tr>
</tbody>
</table>
reliable behavior from organizational members. This function typically is called control (Lawler, 1976). Although the specific methods of control vary a great deal, three general classes of variables are suggested. The first of these is informational. Performance-appraisal systems provide an opportunity to communicate the standards of job performance to job incumbents. The other two of these involve the establishment and communication of contingencies between standards of performance and outcomes. In Table 15.1 the outcomes are clustered in terms of their value or valence to the recipients.

The second cluster of functions of performance-appraisal systems for organizations does not deal directly with individual employees; it deals with information that makes possible the evaluation of organizational systems. Such information is for the establishment and operation of essential organizational practices, but it is not necessary that employees evaluated by the performance appraisal system be aware of the results of their own evaluations in order for the evaluations to be useful for system evaluation purposes.

Turning to individual employees who are evaluated by the performance-appraisal system, three sets of functions are listed. Obviously, in order to meet these functions, the evaluated employee must be aware of his or her evaluation. The first of these functions involves learning the tasks involved in the job. Ideally, performance-appraisal systems with feedback not only communicate to the employees what it is that they are supposed to do but also inform them of the degree to which they have met the standards over the time period covered by the appraisal.

The other two functions are based on the hedonic assumption that employees will want to obtain rewards and avoid punishments. To the extent that performance appraisals convey knowledge about the contingencies between performance of the job and the attainment of rewards and the avoidance of punishments, they serve employees' desires and have the potential of influencing future employee behaviors. I would contend that, to the extent that appraisals aid in learning and provide for the attainment of valued outcomes along with the nonattainment of negative outcomes, they serve the functions of the employees.

**The Functional Role of Appraisal Accuracy.** Note that we mentioned earlier that the underlying assumption regarding performance-appraisal systems is that their fundamental purpose is to provide a valid and reliable measurement of employee job performance. That is to say, the system is a measurement system that should be an accurate reflection of how employees perform their jobs. However, also note that when we ask another basic question, "What do we hope to accomplish for whom with a performance system?," we have generated a set of functions served by the systems without mentioning accuracy. In some cases, this may be because accuracy was simply assumed. Certainly, if organizational decisions are going to be made on the basis of data generated by appraisal systems, we would hope that the data on which the decisions are
based are accurate. Likewise, if individuals are to learn how to do their jobs from the system, accuracy is important. On the other hand, is it necessary that appraisal information be accurate to control employee behavior or to provide the employees with means to obtain rewards and avoid negative sanctions? The answer to this question is less clear. In fact, it is sufficiently unclear to me that it makes me a little uneasy to even ask it, because, in asking it, I feel compelled to consider the alternative that there may be cases when inaccurate appraisals serve functions that seem quite legitimate. At this point, let us tuck this thought away but not forget it.

Performance-Appraisal Options

Performance-appraisal systems vary a great deal in the way in which they are conducted. In order to address the functions served by appraisals, we must first consider how these functions might be met. For purposes of discussion, consider the following general practices.

Coach and Consequences. In most all-inclusive performance-appraisal systems, information is gathered in the system that is communicated to employees and also becomes part of their personnel record. The data in the system are then used for both individual- and system-level decisions. Individual ones involve such outcomes as salary, promotions, job assignments, and dismissals. The performance evaluations are also fed back to the employees to provide them with a basis for maintaining behaviors that were valued and change those that were seen as in need of improvement. Such information is also useful for setting future performance objectives.

Coach. A second common use of performance-appraisal information is directed primarily at the coaching function. In this case, the information is decoupled from administrative functions. It is used primarily to provide feedback for self-evaluation and for setting future performance goals. Two slightly different modifications of the coaching practice address the way in which the data are handled once they have been used for coaching. Most frequently, the data are maintained at the unit supervisor level and not shared with others in the organization. Another alternative is to place the data in a file that can be accessed for making system-level decisions but not individual-level ones. In this case, the data could be used to evaluate training programs or validate a selection system, but the data would not be placed in the employee’s personnel file. These two variations are labeled Coach Only and Coach and Store, respectively.

Research Only. Performance-appraisal data can also be collected only for use regarding organizational-system decisions. The most common of these uses is in validating selection systems. The appraisals serve as the criteria against
which tests, interviews, and other selection materials are evaluated. Data collected for research purposes are also useful for inventorying the performance levels of employees when needs analyses for training are conducted. The common characteristic of the research-only option is that the data are not shared with the employees, and they are not inserted into their personnel records for access regarding individual personnel decisions.

No Performance Appraisal. For completeness, there is one other option regarding performance appraisals. This option is not to do them. Although those of us who are involved with the many facets of human resource management are loathe to consider this option, managers and workers are not so committed to the appraisal process. Therefore, we need to consider this option as we lay out possible responses regarding performance appraisal.

Matching Functions and Options

Table 15.2 combines the functions of performance-appraisal systems with the options. The ‘‘X’’s in the table indicate the functions that are typically believed to be addressed by the particular option that is used.

<table>
<thead>
<tr>
<th>Appraisal Option</th>
<th>Organizational</th>
<th>Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>System Level Evaluations</td>
<td>Control</td>
</tr>
<tr>
<td>Coach &amp; Consequence</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Coach &amp; Store</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Coach Only Research</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>No P. A.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations:
- Crit. = Criteria
- Needs A. = Needs Analysis
- Comm. = Communicate Performance Standards
- Rew. = Administer Rewards
- D & D = Discipline and Dismissals
- Learn. = Learning
- Rew. = Obtain Rewards
- Sanc. = Avoid Negative Sanctions
- Consequence = Consequences
- P. A. = Performance Appraisal
In Table 15.2, the first option, Coach and Consequences, is described as all things to all people. The performance-appraisal system is administered to provide information for the organization to guide system-level decisions regarding the validation of selection procedures, training evaluation, and other practices of human resource management, as well as to control individuals by communicating standards and administering rewards and sanctions. Likewise, the individual is to receive individual feedback and experience the motivational value of contingent rewards and sanctions. If this sounds like too much of a good thing, it is. Although this option is the most frequently attempted option, using appraisals to accomplish all these ends is asking the system to deliver more than it is capable of doing (DeVries, Morrison, Shulman, & Gerlach, 1986). As a result, using the coach and consequences option creates many of the complaints to which we have already alluded among those who use and are affected by the system.

There are two ways to respond to the inability of the Coach and Consequences option to deal with many of its problems. One is to attempt to remove some of the conditions that lead to problems and still try to maintain all the functions. There have been many attempts to do this. Increasing managerial commitment to the system (Banks & Murphy, 1987), creating an attitude of support and constructive feedback (Burke & Wilcox, 1969; Maier, 1958; Solem, 1960), involving ratees more in the system through participation (Mohrman & Lawler, 1983), and numerous other attempts to address variables both within and outside the system (see, for example, Beatty & Bernardin, 1984; Wexley & Latham, 1981) have been used to reduce problems with the system and yet still maintain all desired functions.

A second position is to question whether any performance-appraisal system can be expected to accomplish all the ends outlined in Table 15.2. Such a line of questioning leads to the exploration of conditions that would limit any system that attempts to be all things to all people. The early work at General Electric by Herb Meyer and his colleagues (Meyer, Kay, & French, 1958) recognized the fact that the goal of evaluation for organizational decisions was incompatible with coaching and learning when it was advocated that these two functions be separated. More recently, Longenecker et al.’s (1987) position is that, because organizations are political systems, any performance appraisal tied to individual outcomes will be influenced by political forces to the detriment of rating accuracy.

The preponderance of the data would imply that no system can do all that the coach and consequence option demands (DeVries et al., 1986). Furthermore, when systems try to do it all, the result is likely to be that the data collected by such systems are not likely to serve any of the functions very well. More than likely, the data (ratings) will be distorted (Mohrman & Lawler, 1983).

If a system cannot meet all objectives, an alternative strategy would be to focus on only some objectives. Table 15.2 suggests that there are three limited objectives. The most common of these is the Coach Only option. Here appraisals
are conducted and the information shared in a performance-appraisal review between the appraiser and the appraised. This information is kept confidential between the appraiser and appraisee. Such systems can be valuable for working with employees at the individual level. The limitations that do exist are due to the fact that the employee may not see the appraisal interview as truly independent of the appraisal with consequences. The idea of separating the appraisal interview for consequences from that with only coaching functions, as suggested by Meyer et al. (1958), underplays the fact that the employee is still aware that the same person giving the coaching feedback may also evaluate him or her later (Ilgen & Feldman, 1983). The result of this realization is that the ties to evaluation still influence ratings and responses to them in the coaching setting.

Another coaching option suggested in Table 15.2 is to store the coaching information, but only as criterion data for organization needs, not at the individual level. To my knowledge, this is not frequently done. This may be due, in part, because systems developed for coaching purposes try to avoid any distribution of information beyond the level of the supervisor. Certainly, to do so must involve careful assurances of confidentiality and guarantees that the data would never be used for individual personnel decisions. Another limitation may be that the types of ratings and information most useful for coaching may not be the types that best serve the validation of a selection system or the assessment of training needs. It would seem that both the usage problem and that of the nature of the information could be addressed and would be valuable for increasing the use of the data, whereas at the same time not creating conditions that are likely to conflict in such a way as to lead to distortions of the ratings.

A third option mentioned in Table 15.2 is that of using the performance-appraisal data only for research purposes. There is a great deal of evidence that performance evaluations gathered only for research purposes are of higher quality than those gathered for other purposes (Padgett, 1988). Presumably, the higher quality is due to the fact that the confidentiality of the ratings removes any felt need to distort the data on the part of the raters.

To my knowledge, data gathered for research purposes in the past have only been collected on a single case basis; that is, the research purpose goal of performance evaluations has not been institutionalized. Ratings are not obtained on a regular basis only for research. If the research option is to be considered one of several for performance-appraisal systems and is to be depended on to generate data necessary for criterion development and needs analyses on a regular basis, it would be necessary to institutionalize this process. The advantage of a regular collection of such data under the protection of anonymity is that it would provide a database that is likely to be more accurate and valid than other methods of generating performance data. The disadvantage is that the users of such systems may feel that the time and effort needed for establishing the criterion dimensions, training on the use of the system, and completing the evaluation is not worth it if the data are not used for other purposes; that is,
the disadvantages are likely to be motivational with respect to the raters, not the ratees.

The fourth and final option listed in Table 15.2 is one that those of us involved in developing and implementing performance appraisal systems are loathe to consider—the option of no performance appraisal. For medium-to-large-sized organizations, I believe that it is unwise to have no performance-appraisal system, and I would like to think that this belief is based on more than self-serving biases. On the other hand, if the organization is unwilling to put forth the effort to develop and maintain a good system leaving only a system that is mistrusted and produces ratings of questionable validity, then the option of a no-rating system may be preferable to the one in use, but not to a properly designed and maintained one.

Another option for the lack of use of performance appraisals is to suggest that when, at the organizational level, a performance-appraisal system is in use, there still may be conditions under which performance appraisals should not be used. Let me suggest several such conditions.

In most organizations, there are a number of jobs on which there is very little opportunity for individual differences in job performance. Under such conditions, performance appraisals have little or no value for controlling behavior, nor do they provide much information about job performance over and above what is known from the fact that the person is in attendance and not disrupting the performance of the unit. Furthermore, in such conditions, the employee is likely to gain little or no additional information about his or her performance from the ratings. Therefore, for such jobs, the system is redundant and unnecessary.

In other situations, rather than be redundant, performance-appraisal systems may be disrupting. Consider the situation in which the typical raters of performance (supervisors) do not have the opportunity to observe the performance of subordinates or do not know the jobs well enough to provide the level of evaluation demanded by the performance-appraisal system. Under such conditions, the supervisors lack credibility, and the evaluations that they generate may be more disruptive than valuable. Unless one can change the reporting process or alter the structure of the evaluation process, it is unlikely to be of much value to attempt to fine tune the performance-appraisal instrument or to train the raters.

Work teams provide another challenge for performance-appraisal systems. By their very nature, performance appraisals are individually oriented; they map evaluations onto individuals (Igen, 1988). When team performance becomes the primary unit of analysis, the issue of individual performance in the team often becomes difficult to judge. In addition, it may also be detrimental to emphasize individual differences in these types of work relationships. It is this latter process issue that is often emphasized by those who advocate teamwork and the Japanese norms of team over individual orientations as a reason not to do individual per-
formance appraisals in teams. Although the positions tend to be based more on general reactions than on data, I do think it is premature to dismiss the fact that individually driven performance appraisals, under some types of teamwork structures, may be detrimental.

As one final situational condition that may lead to questioning the use of performance appraisal, let me raise the issue of multiple performance dimensions where some of the dimensions are very concrete and others are much more abstract or difficult to quantify. Under such conditions, the more concrete standards are likely to dominate behavior; that is, because performance can be measured more easily on the concrete dimensions, it will be easier to agree about the extent to which a person has done well or poorly on these dimensions. Thus, if the person has to perform on a number of dimensions, some of which are concrete and others less so, it is likely that the person will "play" to the concrete ones to the detriment of the others. The latter is particularly true if valued rewards are tied to the appraisal system. If successful performance of the job demands spending a great deal of time and effort working on the less concrete dimensions, the existence of a performance-appraisal system may decrease the frequency of the less concrete dimensions.

The conditions just mentioned questioning the use of performance appraisals have all focused on the work setting. Let me suggest that there also may be conditions within the person that suggest that particular persons be excluded from a regularly scheduled performance appraisal. Consider one condition that I contend is not all that rare. This is the condition of an employee who, over time, maintains a relatively constant level of performance; the person has been an outstanding performer for years or even a very satisfactory one. Assume that this employee then performs considerably worse than his or her previous, well-established, pattern of performance. Also assume two other conditions. The first is that there is a readily available reason for the lower performance such as the need to cope with a serious personal problem such as a spouse who is fighting a substance abuse problem or the employee's own physical illness. Second, assume that the employee is well aware of his or her drop in performance over the time period. Under these conditions, if there is a regularly scheduled performance appraisal for the person, it usually is necessary to appraise the person along with all others. The appraiser is then faced with a dilemma; should he or she appraise the person at the level the ratee performed during the time period, or should the appraisal be based on the employee's "normal" performance? Most appraisers would alter the rating in order to protect the employee if the rater believed that the current level of performance was an aberration. However, to do such simply puts into the appraisal system ratings that are less accurate than the rater is capable of producing. Furthermore, if the data are to be used as criteria for organizational decision making, these manipulated data may not serve very well as criteria. I suggest that, rather than leave the rater only the alternative to distort the rating in order to protect an em-
ployee in whom the rater has a great deal of confidence regarding the employee’s future performance, it might be more reasonable to create an evaluation system that allows the rater the option to skip an appraisal cycle for the employee.

The combination of issues raised regarding times when appraisals may not be reasonable suggests the incorporation of the option to omit appraisals when it is likely that the appraisal data will not be accurate or when accurate appraisals may have detrimental effects on individual or unit performance. From the perspective of an individual’s employment with an organization over time, any particular individual in such a system could be missing an appraisal during any particular cycle, either because he or she was on a job where conditions were not good for appraisals or because of some personal characteristics that justified exclusion. To the extent that the appraisal system was used for making individual-level decisions about reward allocations, promotions, and other decisions, procedures would have to be developed to deal with “missing data.” At first glance, such missing data seem disconcerting. However, it should be kept in mind that the missing data exist because the alternative is to get bad data. I suggest that having the option to remove persons from the appraisal process for reasons of the job or of the person himself or herself has the potential for increasing system-level accuracy. The latter should increase the usefulness of performance-appraisal data for both organizational and personal functions. Whether this indeed occurs waits for empirical verification.

RESEARCH IMPLICATIONS OF A FUNCTIONS OPTIONS PERSPECTIVE

Earlier in this chapter, I mentioned that when problems arose with performance-appraisal systems, the initial response was to rely on our psychological roots and search for solutions among those psychological processes that we understood best. In particular, we looked to our knowledge of scaling to provide the answers and relied on our knowledge of training and other methods of influencing human behavior to put our scaling knowledge to work. When that failed, we again turned to psychological zeitgeist to suggest ways to proceed. This led us to focus on the rater as a judge.

If we begin with the point of view that there are several goals to be met with performance-appraisal systems and that there are a limited set of options for meeting these goals, then the search for solutions to problems is broadened. It leads one to ask what critical functions are to be addressed and what mix of options might meet these functions. In the exploration of answers to these questions, the range of relevant issues expands from those of psychological processes. In particular, I believe that the perspective immediately forces one to consider the conditions of the environment in which the systems are to be placed. At this time it is too early to say exactly how this environment should
be construed. However, let me give you one example of the way in which one might approach the problem.

Margaret Padgett (1988), a recent PhD from Michigan State University, looked at the kinds of factors that might lead people to report ratings on the formal appraisal forms that were not in agreement with their best judgments about the person's actual performance. To gather such data was not easy. It required that she put in a great deal of time and effort to gain the confidence of the people so that they would be willing to report discrepancies between how they rated subordinates and how they felt about them. She met with over 100 supervisors individually to discuss the project and gain their commitment to rate a recently appraised employee on a research instrument. The supervisors also provided responses to a number of other variables and gave permission for us to access the rating of one particular employee selected by the researcher.

Without going into great detail, I have presented the model that she tested in her dissertation and the best fit model that resulted from using a LISREL analysis. Figure 15.1 is the predicted model for a number of factors that she believed would affect the extent to which raters would bias their reported performance evaluations in the positive direction compared to what they really believed was the employee's performance level. According to the model, the most salient factor was the extent to which the raters believed they could be honest in their evaluations. This belief was predicted to be a function of the rater's desire to be liked by the ratee, ability to document his or her rating, expected reaction of the ratee to the rating, and the extent to which the employees were

![Diagram of factors affecting rater motivation]

FIG. 15.1. Padgett's (1988) initial model for factors affecting rater motivation to provide accurate performance evaluations.
likely to share their evaluations with others (Appraisal Visibility). Expected reactions, in turn, were expected to be a function of the credibility of the rater, the sharing of the ratings with others, and whether or not there were or were not positive or negative consequences associated with performance evaluations. Finally, the purpose was believed to be associated with the consequences.

Figure 15.2 presents the final structural model. To go into great detail describing this is beyond my purpose here. My point is not so much in the findings as in the approach. What I am suggesting is that research needs to address the contextual factors influencing performance appraisal. Through a better understanding of context along with clearly specifying the functions we want the appraisal to serve, we should be more aware of the fact that no appraisal system is likely to be able to serve all the functions that we have been typically asking it to serve. Once we understand these limits and understand the options available regarding appraisals, perhaps our research will learn more about mix of performance-appraisal options that will meet a more realistic set of functions.

CONCLUSION

Performance appraisals have been with us a long time. Over that time period, those concerned with developing and refining these systems have made a lot of gains. At the same time, criticisms and dissatisfaction have plagued the use of these systems from their inception. It has been argued here that, in spite of the changes, criticisms have changed little in either level (frequency and intensity) or content; complaints that were heard years ago are still echoed today.

Frustrations with some of the limitations with performance appraisals at the end of the 1970s led Landy and Farr (1980) to suggest a change of course in our efforts to improve such systems. This suggestion was heeded, I suspect, far beyond their expectations. The result was a surge of research effort directed at the rating process.

As I survey the performance-appraisal landscape more than 10 years later, I see little reason to be much less frustrated than were Landy and Farr. There has been some progress, as I have mentioned, but the sources of frustration have changed little. Thus, I too would like to suggest that we alter or change, somewhat, our path in pursuit of the perfect appraisal. The suggestions that I may seem radical to those of us, myself included, who have come to be interested in performance appraisals from a broader concern with the psychological processes important at work or from a human resource management view of human behavior that looks to the micro-processes of human knowledge, skills, abilities, and attitudes to structure appraisal systems. I have two suggestions.

First, unlike past corrections that turned from one psychological path to another, in my opinion, needed today is a course less dominated by psychologi-
cal or psychologically-oriented domains. At the theoretical level, the issues involve the nature of the social milieu in which appraisals operate. Practically, the issues are more managerial than technical; that is, the issues may not be so much developing the technical features of the system as they are managing the system so that it can operate effectively. For example, consider the issue of managerial commitment to accurate appraisals. All would agree that such commitment is a necessary condition in order for a performance-appraisal system to operate effectively. However, once managerial commitment has been identified as an important variable, making an appraisal system work in a particular setting is less of a technical issue than a managerial one of establishing policies and practices that will create and maintain commitment. Thus, establishing a new performance-appraisal system will require not only the usual practices of obtaining approval for the system, developing scales, and training users, but it will also require months of "campaigning" for the system prior to its implementation, working to establish ways to reward good appraising behavior, and other efforts far beyond the boundaries of duties and responsibilities of those who developed the appraisal system in the past.

My second suggestion is that the analogy of a path to the perfect appraisal is incorrect. We should be open to the possibility that there is not one path and to the possibility that some paths lead to no appraisals at all. Even worse from our proappraisal bias, I think we should be open to the possibility that perform-
performance appraisals may actually inhibit effective individual or unit performance. If this is the case, then we must be open to the possibility of recommending that appraisals should not be used. From the standpoint of research, our task becomes that of discovering those situations when appraisals are not appropriate as well as identifying ways to make them work better when they are used.

Incorporating the two previous suggestions into the total domain of performance appraisal in organizations creates a two-pronged attack on performance evaluation. The first of these focuses on the technological issues of the past. I have stated all along that we know a great deal about how to construct performance-appraisal scales and train people to use them. We should not forget or ignore this knowledge and should continue to expand it.

Our second focus must be on the context in which appraisals are placed. We must work to understand this context and then work to change or modify the context so that it is possible for the appraisal system to function appropriately. We may also need to fit the appraisal system into the context in a way that the system is most likely to provide meaningful performance-evaluation data in that context. The task of understanding the context is primarily one for the researchers. That of changing and modifying the context goes beyond that of knowledge and expertise to that of power, persuasion, and influence in the organizational system in which the appraisal is to be embedded. The final result will be fitting and modifying the appraisal system and also fitting and modifying the organization in order to create a good match. In that matching process, it should be kept in mind that the modifications of the appraisal system may lead to a complex appraisal system that excludes some jobs and some people from the appraisal process at particular times. The approach is one of discovering boundary conditions for performance appraisals; also it is one that goes beyond the simple discovery of boundaries and then copes with them, suggesting that the boundaries themselves may need to be shaped and molded.

REFERENCES


