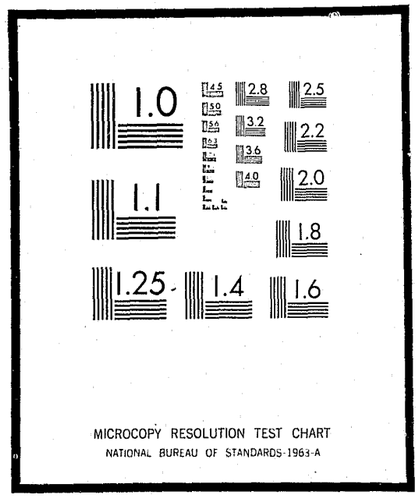

Performance Appraisal for Police Patrol Officers: A Manual for Personnel Officers

Frank J. Landy

and

James L. Farr

Department of Psychology
The Pennsylvania State University
University Park, PA 16802

Prepared under Grant No. 73-NI-99-0036-G from the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, United States Department of Justice.

Preface[a]

Everyone has had the experience of discovering a quotation which perfectly describes his feelings about a particular subject. The opening lines from Winnie the Pooh certainly convey the general concern about performance appraisal as it is administered in most municipal agencies and, in particular, in the municipal law enforcement agency:

> Here is Edward Bear, coming downstairs now, bump, bump, bump, on the back of his head, behind Christopher Robin. It is, as far as he knows, the only way of coming downstairs, but sometimes he feels that there really is another way, if he could only stop bumping for a minute and think of it.

A good deal of every executive's day is spent bumping, and very little is devoted to the thinking necessary for elimination of the bumping. One of the objectives of this manual will be to maximize the thinking regarding performance appraisal. We will first briefly review the evolutionary developments relevant to the evaluation process in police work; we will then identify some of the more common philosophies accompanying performance measurement, pointing out the expected results of each of these philosophies in terms of the decisions that have to be made each day about subordinates; next, we will try to identify current trends in performance assessment; and, finally, we will point out the more common pitfalls to avoid in the actual administration of a performance appraisal plan, some techniques for avoiding these pitfalls, and a general framework for structuring or restructuring an appraisal system.

---

[a]An earlier version of this material appeared in Police Personnel Administration published by the Police Foundation in 1974.

THE STATE OF THE ART

While it would be inappropriate to present a genuine history of performance appraisal in law enforcement agencies in a manual such as this, it is necessary to describe briefly some general trends if we hope to analyze current efforts critically.

Strangely enough, the process and concept of performance evaluation began as an instance in which the cart was leading the horse; it has therefore never quite attained the role which it needs in order to be effective. Historically, development of criteria for success and measurement in law enforcement agencies was an attempt to present some job-related standard for the purpose of determining which hiring tests could distinguish between good and poor applicants. Thus, the emphasis was on initial selection with less attention being given to evaluation of performance after the selection process was over.

No wonder, then, that performance appraisal research failed to keep up with the complexity and "sophistication" of the selection process. While selection methods now include polygraph screenings, psychiatric examinations, and batteries of achievement, aptitude, and personality tests, it is not unusual--in validity studies--to find these predictor scores compared with such criteria as an individual's promotion record or whether or not he remained on the force for a full year. While it would be absurd to argue that these criteria are meaningless, they are less than efficient in describing the total role of a policeman.

A second, slightly more stubborn trend has been the adherence to the global and unitary view of behavior. For lack of a better term, we will refer to this as the "good cop/bad cop" approach. The individual in question is either all good or all bad. There is no concern for strengths and weaknesses. This approach has often been taken because of a belief that the policeman's role is too complex to describe and that the best which might be hoped for is

an overall rating or ranking on something like "quality of work" or "overall suitability". While such a view of behavior might be acceptable for certain limited purposes, it is a totally inefficient use of the evaluation process for providing input to other components of the personnel decision process.

Beginning in the 1940's, attempts were made to capture the complexity of the policeman's role as well as to standardize the process. It was hoped that this would be accomplished through the use of graphic rating scales. Since these scales are still widely used, a separate discussion of them occurs later.

The studies on performance appraisal between 1917 and 1961 show relatively little progress. In the 1960's, however, the subject finally began receiving the attention it deserves. Researchers have been more willing to view the evaluation process as something more than a necessary evil. They have come to view performance appraisal as possibly one of the most important components in an integrated personnel system. As a result of this increased attention, significant studies have been made. But before one can make use of the technical advances which have been made in the field of performance evaluation, it is necessary to examine the philosophies which underlie the evaluation process itself.

PHILOSOPHIES ACCOMPANYING PERFORMANCE EVALUATION SYSTEMS

All police agencies will almost uniformly state that the purposes of performance appraisals are to (a) standardize the nature of the personnel decision-making process so that the rights of the job incumbent are fully represented, (b) assure the public that the agency representatives are fully qualified to carry out their assigned duties, and (c) give the job incumbent the necessary information with which to modify his own behaviors (that is, maintain behaviors which are appropriate from the organization's standpoint and eliminate those which are inappropriate). On a more procedural level police agencies might

also claims specific objectives for the evaluation process relating to other aspects of personnel management: that is, to provide a basis for determining recruiting policies, selection policies, specialized duty assignments, or even pay differentials.

In spite of their general acceptance, such reasons for appraisal do not appear with the same consistency among the individuals responsible for meeting the overall objectives of the police agency. There are a number of executives whose evaluation of subordinates would bear little resemblance to such stated policies. Some of the more common philosophies held may be expressed as follows:

(a) The department says it should be done "x" times a year whether he needs it or not!

This kind of mechanical philosophy (if it deserves being called a philosophy!) is probably the most common in larger agencies. It seems to be positively related to the number of subordinates who must be evaluated. There is little or no concern for the ways in which the information will be used. Evaluation is viewed as a procedure which must be completed to keep the personnel officer off of the supervisor's back!

(b) If we only had a chance to sit down and talk it out, everything would be all right.

This is best characterized as a "human relations" philosophy in which the lion's share of the time is spent with the poor performers. It is felt that the problem is simply one of communication. The supervisor feels that all problems can be solved if the subordinate feels that someone "cares".

(c) It is the only way to justify a personnel decision which has been or will be made.

In this approach the decision has often been made before-the-fact. It is obvious to everyone that the individual in question "has what it takes" or is "a lemon". The specific evaluation information is frosting on the cake. It is used to justify the decision to the individual or others related to the decision.

(d) There are certain minimum standards for performance in this department, and if an individual cannot meet these standards, the sooner we weed him out the better.

This view assumes that the major responsibility for poor performance is borne by the selection process. The task of the superior is to provide information to those administering the testing program that might enable them to change their selection procedures and yield a higher quality officer.

(e) Every individual has both strengths and weaknesses; every individual seeks information necessary to maintain his strengths and eliminate his weaknesses.

This philosophy defines the role of the evaluator as that of a collaborator with the job incumbent. Performance evaluation is a joint venture meant to satisfy both the incumbent and the organization.

While these philosophies are not meant to exhaust the domain of personal attitudes in performance appraisal, they are representative of some of the more common value systems operating. Some of the positive and negative effects of each of these philosophies, both in terms of the organization and the job incumbent, are worthy of brief exposition.

1. The fixed interval philosophy

This philosophy has a rather dramatic effect on individual behavior. The incumbent sees the process as the most important focal point, not the behaviors preceding the process. Since he knows that the evaluation is done at regular intervals, and more importantly, he knows what those intervals are, he will concentrate his energy in the period of time immediately preceding the evaluation. He reasonably assumes that his superior considers the evaluation process a necessary evil and that the superior will not worry about the evaluation until the last minute.

A more serious problem which accompanies this philosophy is the possibility that the employee is evaluated without his knowledge. This is easily cured by

having the individual sign his performance evaluation form, a common practice in police agencies. In private industry, it is not uncommon to find employees who are dissatisfied with management because they feel that they are being ignored. Often an individual will complain that no one has evaluated his or her performance in two or more years when, in fact, the records show that the individual has been evaluated several times in that period.

2. The human relations philosophy

This philosophy assumes that a subordinate can be "loved" into efficient behavior. It assumes that all people are capable of recognizing deficiencies and is usually directed towards those having performance problems. There are several reservations about this philosophy and its assumptions. It is not unusual to hear supervisors complain that: "You just can't get through to that guy. No matter how friendly I am toward him, he continues to goof off." Such a complaint could mean either of two things: (1) the individual does not value the friendship of the supervisor, or (2) he values it very much and sees that the more he goofs off, the more he can expect friendly overtures from the supervisor.

The supervisor is on the right track in assuming that it is necessary to interact with the individual and avoid the temptation to make the evaluation process an impersonal routine. He is on the wrong track in assuming that it is unnecessary to point out specific weaknesses with suggestions for improvement. There may be a curious "positive" outcome of this kind of philosophy: Employees who find human-relations-type supervisors either a bore or a chore and who feel uncomfortable in these "problem-solving" sessions might improve their performance (assuming they know what needs improvement and how to improve it) simply to avoid lengthy meetings with their supervisor!

The organizational outcomes of such a philosophy are, however, rather serious. Energy is directed toward the lower end of the performance distri-

bution. Good performers are assumed to have no problems and need no encouragement. It fails to recognize that only an unusual individual can maintain high quality performance patterns for long periods without some feedback.

3. Rubber-stamp philosophy

When such a philosophy is used, the evaluation process quickly loses credibility. The employee is unwilling to accept the process, because he realizes that the information is being shaped to fit a decision rather than the other way around. The formal evaluation process is a waste of time and energy for the organization, since information other than that included in the performance evaluation scheme is considered sufficient for making a personnel decision. This "information" is generally an overall favorable or unfavorable impression of the incumbent. The evaluation is intended, in this case, to provide justification for the impression.

In practice, evaluation becomes an art rather than a science, where the evaluator or "artist" paints a favorable or unfavorable picture with the available data. Every agency has at least one individual who says: "I can tell whether or not he is a good guy just by talking to him." That may be completely true. It is the responsibility of the agency to have this individual get his methods down on paper, to communicate his techniques to others, and if they work, to make them a formal part of the evaluation process.

4. The selection philosophy

The assumptions accompanying this philosophy are rather simple: (1) the selection process should provide people who are, by definition good performers, (2) people are incapable of changing behavior patterns, and (3) the work role of the incumbent is constant.

It is becoming obvious as more research is done in selection procedures for law enforcement agencies that a demonstrably good job of selection is not being done. There is no doubt that the process can be and is currently being

improved. Nevertheless, we are nowhere near the point at which the burden described above can be placed on the selection process. It may very well be that performance evaluation is the primary tool for making the kind of decision which should be made in the selection phase. There is also the cost/benefit or payoff concept coming into play. From the point of view of the organization, a considerable amount of time and energy has already been spent on the individual. Simply from a pragmatic standpoint, it would be more efficient to try to shape the individual into an acceptable performer and capitalize on the organization's initial investment.

The generalization regarding ability to change is simply wrong. Given appropriate feedback (the primary intention of the performance evaluation) and appropriate incentives, experience shows that almost all individuals are capable of changing behavior patterns.

The third assumption contends, by implication, that the selection procedure automatically changes at the same rate and in the same direction as the work role. This is a rather optimistic and perhaps even naive view. Assuming that the selection process is flexible enough to make changes as a function of work-role changes, there is always a lag. New job analyses will have to be done, new norms established, new selection devices established, pretested, and validated; in short, it sometimes takes years for the selection process to catch up with work-role changes. This means that, although the selection process might have provided the supervisor with an individual having a high probability of job success, it should not have to bear the burden of post-hoc and work-role changes.

5. The individual differences philosophy

This is the philosophy which the authors feel is most supportable from experience and research. It specifies that performance evaluation is a problem-solving process involving collaborative energy expenditure on the part of the

evaluator and employee. It addresses both strengths and weaknesses and, consequently, good and poor performers. The evaluator is seen as a cohort rather than inquisitor. The credibility and acceptance of the process by the incumbent is increased and the organization is capable of making rather immediate changes in the behavior of individuals to accompany redefinitions of work roles. An added by-product is the collection of information which might be utilized in making lateral job changes such as special duty assignments and vertical job changes such as promotions.

RELATION TO PERSONNEL DECISIONS

At this point it is well to take a look at the relationship between performance appraisal and specific personnel decisions which have to be made.

Recruiting

The general problem associated with the recruiting process is one of getting people to apply when they are needed. Attention is directed toward advertising and other incentive schemes. Unfortunately, almost all recruitment research centers around whether TV advertising is more effective than newspaper advertising, whether personal visits by recruiting teams to shopping centers are effective recruiting tools, etc. These studies seem to be concerned only with the number of applicants attracted rather than the quality of the person recruited or his relevance to the job role.

An obvious concern for the personnel officer of law enforcement agencies would be the quality of officer yielded by varying recruiting methods. The only way of determining the answer to such a question would be to measure the performance of individuals to see if a group recruited by one method displays a different pattern of strengths and weaknesses than other groups.

It is conceivable that as work-role demands change recruiting practices should display accompanying changes. A prime example of this kind of concern is the present investigation of the adequacy of women for patrol work and the reasonableness of height and weight requirements. If it turns out that individuals under the height and weight minimums or women can perform effectively in all or parts of the work role (as determined by a system of performance appraisal), then the recruiting system must change in accordance with that new information. It is obvious that to answer such an important question as this an adequate system of performance appraisal is essential.

Selection

The technical term for the relationship between predictor performance (for example, tests) and job performance (for example, supervisory ratings) is validity. It is logically impossible for an agency consistently to select appropriate trainees if they do not know what dimensions of performance are necessary for satisfying the demands of the particular work role. Validity of predictor devices cannot be determined until adequate criteria exist. Such criteria are really measures of performance.

Training

The selection literature is replete with validity studies demonstrating the relationship between selection devices and training or academy performance. It is extremely difficult to identify studies which demonstrate a relationship either between selection devices and on-the-job performance or between academy scores and on-the-job performance. One of the implications of this condition is that there seems to be little or no relationship between training programs and on-the-job performance. Even if a relationship were acknowledged but considered lacking in demonstration, it would take a very unusual training director to develop a consistent and coherent training program in the absence of infor-

mation that relates the program to on-the-job performance. Even then, the training program would undoubtedly improve were such information available.

Probationary Periods

While the probationary period is normally viewed as an extension of selection and training policies, there are certain unique considerations here which warrant attention. From an organizational point of view, this period on the job permits gathering work samples or behavioral measurements rather than academy-class grades or test scores. In addition, the simulations which may have been programmed into the training experience are no longer simulations but real life experiences--stress is now STRESS, danger is now DANGER, compassion is now COMPASSION. If an efficient system can be set up for measuring reactions in these kinds of situations, there is a good chance of maintaining appropriate behaviors and correcting inappropriate behaviors before they become habit. From the standpoint of the officer on probation, this is the most important time for feedback. Everything he does is new and has no necessary organizational value sign attached to it. This is the opportune time to supply performance feedback.

Horizontal Job Changes

In many respects horizontal job changes (e.g., special duty assignments) and vertical job changes (promotions) are similar. In both instances, an attempt is made to match an individual with a work role.

The phrase "special duty assignment" implies that there is something unique about the activity in which the job incumbent will engage. In practice there are many unique aspects such as the kind of person the incumbent is likely to come in contact with, the type of crime being considered, the particular skills which are brought to bear in the situation, and the like. Interestingly enough, if jobs which are not considered "special duty assignments" were analyzed, there

would likely be many dimensions of performance similar to those of the "special duty assignment". They might differ in their importance or the frequency with which they come into play, but there would be overlap. Whatever the differences, they require specific identification.

This is where the adequacy of the evaluation system becomes essential. It is not sufficient to specify that the individual's "quality of work" is satisfactory and, therefore, he should do well in the new position. It is obvious that the nature of the work roles are different and that a general, global performance dimension like quality or quantity of work is seriously deficient information on which to base a transfer decision. Only when the tasks represented in each of the positions are identified, when the individual in question is appropriately evaluated on each of these dimensions, and when individual characteristics are compared with work role requirements, can a successful transfer be effectively assured. This is a rather oblique way of making a rather simple point--the major personnel function of the organization may be one of placement rather than selection. The sequence of activities is carried out literally every day in most agencies in one form or another. Unfortunately, it is carried out as an "artistic" rather than a scientific endeavor with all of the accompanying pitfalls of artistic personnel systems.

Vertical Job Changes

All of the comments directed toward horizontal job changes apply equally well to vertical job changes. There is, in addition, a painfully obvious consequence of ignoring performance evaluation information in making promotions and using promotions strictly as a reward system without any concern for the strengths and weaknesses of the individual in question. When an individual has done an outstanding job for a period of time, it is natural and appropriate for the organization to reward him. Curiously enough, the only way an individual can be substantially and differentially rewarded in a good number of orga-

nizational settings is with a promotion. Promotions usually are accompanied by increased status, increased responsibilities, increased freedom of action, more money, possibly better fringe benefits, etc. It is often difficult to provide the employee with these kinds of differential rewards without a change in job title and usually a vertical change. Here again, the need for analysis is evident--analysis to make certain that promotees possess the new qualifications needed in the new job and consideration to determine whether promotion is the most rational reward in any given instance.

Salary Increases

Money is probably one of the more potent rewards in many occupational work roles. Law enforcement agencies are certainly no exception. Money is a generalized reinforcer--it can mean bread on the table to one person, it can mean status to a second; a third might conceive of it as the only means by which the organization can formally recognize his efforts. In short, it has the potential for satisfying many diverse needs. As such, it has the power to influence work-related behavior.

When money is applied as a reward in an organizational setting, it also performs the function of giving the individual information about his performance. An increase says, in effect, "you have been doing a good job; keep up the good work." More importantly, it implies that "everything which you have done in the past x months was good." Conversely, a smaller increase or no increase means "you have been doing a bad job in the last x months." However, money by itself, either its presence or conspicuous absence, gives the individual no information which he can use to maintain specific strengths or eliminate specific weaknesses.

Therefore, an efficient use of money as a reward implies that (a) increases should be based on evidence of behavior that the agency wants to maintain in the individual and (b) the individual should be made aware of his behaviors which

led to the salary decision. Neither of these conditions can be satisfied when decisions are made on the basis of inadequate and/or inappropriate evaluation systems.

Disciplinary Actions

Many of the comments made in the section dealing with salary increases are appropriate here. In particular, the feedback of information is essential. If specific performance information accompanied by specific action plans is presented to the incumbent, there is a much better chance that inappropriate patterns will disappear. There is also a lower probability that formal disciplinary action will be necessary. That is not to say that situations of graft and corruption, for example, could be eliminated if only the chief bothered to tell the individual that graft was wrong!

There is another problem which presents itself when an agency depends primarily on disciplinary actions to control behavior. Punitive actions generally tell an individual what he cannot do but seldom tell him what he can do. In addition, many punitive systems provide a challenge to the incumbents to try and outwit the system, to see how much they can get away with before being "burned".

PROBLEMS IN GATHERING PERFORMANCE INFORMATION

Even when convinced of the value of a coherent and systematic performance appraisal scheme, an organization has additional obstacles to overcome, generally relating to the particular devices used to gather the performance information. It seems useful therefore to describe some of the more popular techniques currently in use as well as problems associated with them.

The Matter of Objectivity

Unfortunately, in many occupations it is not possible to deal with objective data, such as units of production or error rates. Law enforcement work

is certainly one of them. Such objective data as exist in police performance usually take one or more of the following forms: (1) separation from the force, (2) performance in recruit training program, (3) promotion, (4) consideration for preferred assignments, (5) commendations, (6) charges of misconduct, (7) vehicle accidents, (8) absenteeism record, (9) accident record, and (10) number of arrests in various categories. The biggest problem with the use of these kinds of measures is that they are seldom under exclusive control of the individual being evaluated. An additional problem is that many of the measures (such as commendations and disciplinary actions) apply to a very small subset of the organization; a corollary to this point is that the measures are more directed toward the extremes of behavior than commonplace behavior. The measures or events cited may be the peripheral by-products of what a policeman does but in no way do they fully describe his role in functional terms. Think of how strange it would sound in counseling a subordinate to say, "Get more commendations" or "Get a promotion". These are not the ways in which one thinks of the work role.

Subjective Standards of Performance

Because of the need for a more functional description of the work role of a policeman, information in the form of subjective or judgmental data has become very important in the performance evaluation process.

Rating scales. The most popular form of subjective evaluation is the rating scale. The graphic character of such scales is intended to permit a supervisor to specify a particular level of behavior or performance for each of a number of factors (quality, quantity, conduct, etc.) in a person's work. Horizontally on one line he can choose the most appropriate description or degree fitting the employee's case with regard to that factor. From these, a profile or summary is usually drawn.

There are three major dimensions of graphic rating scales: (1) the degree to which the meaning of the response categories is defined, (2) the degree to which the person interpreting the ratings (for example, a higher official or the personnel officer) can tell what response was intended by the rater, and (3) the degree to which the performance dimension rated is defined for the rater (see, for example, Robert M. Guion, Personnel Testing, McGraw-Hill Book Company, 1965).

"Anchoring" is the process whereby the meaning of each of the response categories is defined. Just as a ruler is marked off in inches or centimeters, the rating scale must be marked off in units of some type if the rater is to make discriminations among people which are of any value. Whether these units are descriptions of different behaviors or numerical values or both, they leave a lot of interpretation to the supervisor. If numerical values are used, they generally are more useful when fairly specific descriptive statements accompany them.

The second major characteristic is the degree to which the person interpreting the ratings can tell what response was intended. Another name for this characteristic might be "response ambiguity". The qualifying conditions affecting the scaling expressions apply equally to the reader of the rating as to the rater.

The third dimension is the degree to which the performance factor rated is defined for the rater. Consider all the possible interpretations which raters might apply to the simple term "quality of work". This allows for the possibility that a set of ratings on employees might be worthless because they are based on variable interpretations by the raters of the factor to be considered.

Obviously the ambiguity of the rating scale format can lead to errors in ratings which render them useless. These errors can be placed into three major categories: leniency errors, halo errors, and central-tendency errors.

Leniency errors. It is not uncommon to run across raters who may be described as unusually "harsh" or unusually "easy" in their ratings. These extremes usually come about because the rater is applying his own standards to the rating scale. For example, the anchor description "excellent" may have two completely different meanings to two different raters. Two suggestions have been offered to eliminate these errors: (1) a forced distribution in which the rater is required to allocate his ratings in a pre-determined distribution so that a given percentage of the ratees must fall in the top category, a given percentage in the next category, etc., and (2) reduce the degree of ambiguity of the scales by improving the definition of the dimension and the nature of the anchors. The second suggestion is somewhat easier to implement than the first. The forced distribution implies some clear knowledge of what the distribution should look like or how the trait should be distributed among the employees. Such knowledge is not easily acquired.

Halo errors. The term "halo" is meant to imply that there is a general aura which surrounds all of the judgments on separate factors which are made about a given individual. The rater has a general favorable or unfavorable impression of the subordinate. He then proceeds to assign ratings which are consistent with that impression. Halo errors may also come into play when a rater feels that a particular factor of performance is extremely important and then assigns ratings on the other factors which are consistent with his rating on the most important one. One method that has been suggested for solution of this problem is to rate all of the employees on one factor, then all of them on another, until all have been rated on all factors. The hope is that the rater will be able to distinguish between factors if he is forced to consider all of the ratees on one at a time. It might also be suggested, of course, that if the factors were clearly defined and anchored, the raters would be less likely to depend on overall impressions.

Central-tendency errors. Central-tendency errors are characterized by unwillingness on the part of the rater to assign extreme ratings--either extremely high or extremely low. This error has a potentially serious effect on attempts to establish empirical validity for selection devices. Since the range of variability is restricted on the criterion, the value of the obtained validity coefficient may be drastically reduced. Once again, the forced distribution method has been suggested as a way of eliminating central-tendency errors.

Most of the suggestions for dealing with rating errors deal with changing the form of the rating process. Others have suggested that a supplementary approach for the correction of these errors is the education of the raters, that is clarifying to the raters the existence and meaning of these tendencies. The hope is that their effect will be reduced when people are made aware of them. It is strongly recommended that raters receive proper training regarding the how's and why's of performance appraisal. It is especially important that new sergeants and other recently promoted officers be introduced to the process and issues. Refresher courses are also necessary so that bad rating habits do not develop.

Alternatives to Rating Scales

Attempts have been made to minimize the effects of halo, leniency, and central-tendency errors by using alternative evaluation schemes. Two of the most popular of these alternatives have been the employee comparison methods and the forced-choice checklist format.

Employee comparison methods. The two most widely used employee comparison methods are the ranking procedure and the pair-comparison procedure. In the ranking procedure, the rater simply arranges the employees in rank order from high to low on a given dimension of performance. This process can be simplified for the rater somewhat by using an alternating procedure in which he picks out the best and the worst man, sets them aside, picks out the best and worst from the remaining set of names, and so on until all the employees are ranked. One

of the problems in the use of this method is the nature of the data which result from the ranking. It is of little value to know that an individual is ranked 3rd of 30 on some performance dimension unless one has some notion of whether or not that represents adequate, good, or bad performance. Rankings do not give that information. Another criticism of the ranking methods currently in use is that the ranking is usually done on one overall suitability category. When more than one factor is used, ranking may suffer from halo error to the same degree as graphic scales.

As an alternative, instead of simply ranking the incumbents on an absolute basis, or using an alternating ranking procedure, each individual may be compared with every other individual in the group. The number of resulting pairs can be calculated from the formula "n(n-1)/2" where "n" is the number of people to be evaluated. If there are 10 people to be evaluated, there will be 45 pairs. The task of the evaluator is to choose the better of each pair. As in the ranking procedures, in most instances a global dimension of suitability is used as the criterion for making the choice.

Employee comparison methods have the following weaknesses: (1) comparability of evaluations across groups or locations is often lacking, (2) the intricate statistical assumptions which should be met in such a method are often ignored, and (3) the use of one global factor yields information which is unsuitable for making specific personnel decisions such as job change. Such attempts to correct the deficiencies of standard graphic rating scales therefore may simply introduce different kinds of error rather than eliminate error.

Forced-choice formats. In the forced-choice format the rater is presented with sets of items, usually four per set, which he is to use to describe the subordinate. He is instructed to choose the statement(s) most or least descriptive of the person being evaluated. The statements have been carefully chosen for the list as a function of two properties: (1) favorability--the degree to which the item is used as a description for all people, regardless

of whether or not they are good performers, and (2) power of discrimination--the degree to which the item actually distinguishes between good and poor performers. While it is hoped that this method reduces the amount of rater error present in the evaluations, or at least indicates those raters who bias the evaluations, forced-choice formats are rarely designed to yield specific behavioral information but rather yield global estimates of things like "general suitability".

The Ohio State Highway Patrol, with the assistance of the Ohio State University, has developed and used a forced-choice rating format for a period of 14 years. During that time the method has gradually evolved to a point where the Patrol is able to determine the traditional pile-up of performance scores at the high end of the scale, eliminate artifical between-district differences, and decrease the effect of personal bias on the part of the rater. Also, one of the major complaints against forced-choice scales (that is, that the result is usually one overall score) has been addressed. In one variation of the form, four separate dimensions of behavior can be identified. In addition to the traditional forced-choice format, where the rater is asked to choose the two statements most descriptive of the employee, Ohio has incorporated a forced-ranking procedure. In this format the rater is required to rate each of four statements on the basis of how well it describes the ratee.

While some problems still exist in determining exact behavioral descriptions of various points along the particular factor in question, this general line of research seems to be proving fruitful. Other agencies are in the process of developing forced-choice inventories.

RECENT DEVELOPMENTS IN PERFORMANCE APPRAISAL

Devices

There has been a major development in the past decade relating to the devices which are used to gather performance information. This is a restructuring of the traditional graphic rating scale into a "behaviorally anchored" scale.

A behaviorally anchored scale is a much more carefully designed format in which: (1) the anchors are very specific, (2) the anchors are written in a language understandable to the rater, and (3) the factors to be rated are well defined. In addition, the instructions which usually accompany these scales make it easier for the rater to specify exactly what value he had in mind when assigning his ratings.

The strength of this kind of scale results from the method used for scale construction. An outline of the method used to develop the scales will make this strength more obvious:

1. A sample of job incumbents is gathered in a conference setting to define all of the independent performance factors in their positions. The conference results in a consensus decision about major factors and their definitions.
2. A second sample goes through the same procedure.
3. Results of the two conferences are compared and factors which appeared in both conferences are retained.
4. A third sample of incumbents is requested to supply preliminary "anchors" in the form of high, medium, and low examples of performance on each of the independent factors.
5. A fourth sample is provided with a list of all examples written by the third group as well as a list of the major factors and their definitions. The task of this group is to try and place the examples in the category for which they were written.
6. A fifth group is requested to attach specific values to the items or examples which survived the previous step.
7. Those items which receive essentially the same scale values from Group 5 form the final scales.
8. Interrater reliabilities are determined by having individuals rated by two supervisors who are familiar with their behavior.

While the procedure is time-consuming and cumbersome, there is some evidence that the resulting scales are reliable and usable. In terms of their purposes, Steps 1, 2, and 3 are intended to insure that the dimensions are general enough to apply in many different situations and are not the product of one unique individual or situation. Step 4 is intended to provide anchors which are behaviorally rather than evaluatively based. In addition, the anchors are provided in the language of the people who will ultimately use them. Step 5 insures that each example is actually related to the category in which it appears rather than some other category. If there is some disagreement among judges about exactly what kind of behavior is being described, this step will identify that lack of agreement and the item will be dropped or rewritten. Step 6 insures that the final items or anchors actually distinguish between high and low performance. If there is a wide variation in the scale values assigned to any given item, it is an indication that the item is not a particularly good one for distinguishing between good and poor performance. Step 8 provides information related to the consistency of rating among raters.

A major effort which is being made at the present time to develop behaviorally anchored rating scales for use in a wide variety of law enforcement agencies is the present project which concentrates primarily on patrolmen and attempts to set up a model for personnel decisions with performance evaluation as the keystone. See Landy and Farr (1975) for details of the development of these scales. The results of this project are encouraging and may provide the basic approaches and devices necessary for the full utilization of performance appraisal information in making personnel decisions.

Strategies

Developments in devices are accompanied by developments in general strategies, some of which have already been alluded to in the foregoing discussion.

They are: (1) peer evaluation, that is, appraisal by persons of the same rank, and (2) involvement of incumbents and supervisors in the development of a performance appraisal system.

Both formal and informal research is being done on the use of peer evaluations. There is a lot to be said for this approach. It is conceivable that some behaviors are best described by peers rather than supervisors. Supervisors often have an inadequate opportunity to observe certain behaviors. There are other behaviors which are observed both by supervisors and peers. Peer evaluation adds both unique information (those behaviors which cannot be reliably reported by supervisors) and confirmatory information (those behaviors which can be reliably described by both supervisors and peers). In addition, peer evaluation would seem to fit nicely with the increased sense of personal professionalism being manifested in law enforcement agencies.

A second strategical development which accompanies the development in devices is the increasing use of incumbents and supervisors to help construct the evaluation system. They usually have the opportunity to react to a finished product rather than to help in its development. In addition to the valuable ideas which these individuals may have to contribute, this may provide an excellent opportunity for training supervisors in the use of evaluation devices and employees in the value of a comprehensive appraisal system.

OTHER CONSIDERATIONS IN THE EVALUATION PROCESS

Frequency of Evaluation

The frequency of evaluation obviously refers to the frequency with which a supervisor formally records data relevant to the behavior of subordinates for departmental use. In actuality, seldom does a day go by in which he does not evaluate the performance of subordinates. But, more often than not, these are mentally noted rather than recorded observations. When the time comes for

collating this information, the only thing left is a general impression. In these circumstances the official is inclined to remember the things which happened most recently. Consequently, feedback is structured around recent events, and the incumbent is reinforced in his belief that the most important time period for evaluation is that period immediately preceding the formal process.

It is obviously impossible to conduct formal, full-blown appraisals every week. It is not, however, impossible to keep a diary of incidents, both good and bad, which will help provide feedback when the time comes for the formal evaluation. The person being evaluated will then, more appropriately, believe that acceptable performance is not something to consider only during the two weeks preceding the formal evaluation but throughout the entire period from one evaluation to the next.

There is little reason for placing faith in the power of performance review for changing or maintaining behavior patterns if it is not done fairly frequently. Since the probationary period is so crucial for establishing appropriate behavior patterns, formal evaluation and feedback should be done most frequently during this period, perhaps every month or six weeks. At the opposite extreme, for an employee who has served well for many years in the same assignment and with the same expectations, it is pointless to insist on frequent formal evaluations which may result in tiresome and embarrassing repetition of the obvious--unless there has been a notable change in job demands or in performance. In fact, in the case of very long-term employees, fixed periodic appraisals may be dispensed with entirely--aside from the need to reassure them informally that their contributions are still valued.

Discussion of Performance Evaluation with Incumbent

The most carefully constructed performance appraisal system is a waste of time and effort if the information gathered is poorly used. If performance

appraisal is seen as a method of strengthening the behavior of the individuals on the force, it is imperative that these individuals receive feedback about their performance. Supervisors are often reluctant to discuss shortcomings with subordinates. This is one of the major reasons for the commonplace "sign-off" procedure in which the person evaluated must sign his completed evaluation form. Often he perceives his role as one of an inquisitor rather than advisor. This may be a function of the information which is at his disposal. If the evaluator is given information which enables him to address specific behaviors rather than global undefined traits, he may very well feel like a collaborator in the process of improving performance instead of an adversary.

The way in which the subordinate approaches feedback also influences the value of the information he receives. If a supervisor takes acceptable behavior for granted and concentrates only on unacceptable behavior, the subordinate will quickly learn to perceive the feedback sessions as aversive or punishing. In such instances it is not unusual for the subordinate to spend most of his time waiting for a chance to tell his superior what he has done right while attempting to defend what the superior considers to be inappropriate behavior patterns. In this case both parties are talking, neither is listening. We can expect little in the way of positive change from such an interaction. It is essential that the individual be given both positive and negative feedback in as specific a behavioral framework as possible. He should be given suggestions or strategies for improving in deficient categories, not an arbitary ultimatum to "shape up". An appropriately conducted performance review session has an impact on the satisfaction and motivation of the individual as well. This is a time when some of the more abstract rewards, such as praise recognition, that are essential for developing a sense of professional pride in one's work can be applied.

The Combination of Objective and Subjective Data

While it would be convenient to have a simple formula for the combination of objective data (for example, misconduct charges or commendations) and subjective data (such as ratings on various factors of performance), it is probably inappropriate due to the nature of the data involved. Most data of the objective type are not really behaviors but rather the manifestations of behavior.

To give a concrete example, if an individual has had a large number of vehicle accidents over a period of time, it would be of little use simply to tell him "have fewer accidents". One must be able to identify the pattern of behaviors that may contribute to accidents; the problem may really be that the individual is engaging in unsafe practices. In such a case standard objective data might be used as first-level diagnostic information rather than as descriptive behavioral information. In other words, this information could be used to determine which areas of the subjective data bear closer examination. Objective data need not be ignored, but rather they may fill a unique role, a role quite different from that of subjective data.

A variation of this theme might be to look at the empirical relationship between patterns of subjective data and particular pieces of objective data. For example, can specific patterns of performance be identified which are related to commendations or misconduct charges? If one can identify these patterns in terms of the behavior in which the individual engages, one can systematically set about achieving organizational goals through the modification of that behavior.

STRUCTURING A PERFORMANCE EVALUATION SYSTEM

Before specific action plans are outlined for building an evaluation system, a few warnings are in order. The first relates to time expenditure.

If a system is to be both logically grounded and workable, appropriate time and energy will have to be devoted to its development. It will take a considerable number of man-hours as well as a good deal of technical expertise. Temptations in the direction of "quick and dirty" systems should be avoided. They will probably do more harm than good. In all probability there will be a strong positive relationship between the care devoted to the construction of a system and its ultimate value.

This leads to a second warning. The determination to use someone else's system should be based on the technical, philosophical, and practical foundations of the system, not on who is currently using it. The system may work well in one agency but be totally unsuited for another situation. The only way of determining its potential is to examine its development, the situation in which it is now being used, and its intended purpose, as well as the new situation itself. It is very tempting to adopt someone else's system lock, stock, and barrel, but the borrowing process may simply incorporate errors as a function of the new application.

With the warnings out of the way, the general procedure for developing a performance appraisal system may be summarized with these steps:

1. Determine why an appraisal system is desired in the first place. Is the information to be used solely for traditional personnel decisions or for counseling purposes? Are behaviors to be related to a broad range of traditional and non-traditional personnel decisions or just to one decision?

2. Conduct a job analysis of the positions which fall within the scope of the performance appraisal system. There are several different approaches which might be considered in the choice of a job analysis scheme, such as task-oriented vs. worker-oriented analyses. While it is inappropriate to describe these various methods in any detail, it should be pointed out that one must be able to identify the major components of a particular work role before performance distinctions among people can be made.

3. Evaluate the various devices which are available for measuring performance on the various factors identified in the job analysis. There are many different variations of devices available. The decision as to which should be used will most likely be based on the nature of the people providing information (for example, peers, supervisors, citizens) and the use to which this information is to be put.

4. Provide opportunities for input from supervisors and incumbents in the developmental phases of the system. Regardless of how technically and logically sound a system is, it is useless if it cannot be implemented at operating levels. One of the best ways to get support at those operating levels is to have superiors and employees invest something of themselves in the system. This can be achieved by working with them in the development of the system. When the goals of the agency become the goals of the police officers, the success of the system becomes rewarding in and of itself. As a consequence, they will be more likely to expend energy in making it work.

5. Determine the properties of the evaluation devices which are chosen or have developed. Are they reliable? Do they provide the information hoped for? Can they be used with any degree of efficiency by supervisors, peers, and others?

6. Set up a mechanism for information gathering and feedback. Insure that the system will be used as it was intended. Specify the frequency of appraisal and the nature and frequency of feedback sessions.

7. Determine a "probationary period" for the system. Do not assume that the system developed is good forevermore. Designate a specific time for the evaluation of the system itself, and get evaluative information about it from all those affected after it has been in use for a given period of time. Be prepared to make modifications if it is not living up to expectations.

8. Set up a mechanism for keeping up to date on developments in the field of performance appraisal. Designate someone to keep up with the professional

literature; get on appropriate mailing lists; set up working relationships with other agencies for the free exchange of ideas and strategies.

While this postulating may seem a little vague in isolation, most of these suggestions have been presented in greater detail in earlier sections of the chapter. No one can provide a formula for developing the perfect system; if it is to be worthwhile it must respond to an organization's unique circumstances. Nevertheless, based on logic and information, it is possible to design a system that provides the necessary amount and kind of information with which to make the broad range of daily personnel decisions regularly confronting police administrators.

SOME FINAL POINTS OF IMPORTANCE

Here we wish to repeat a few points mentioned earlier as well as introduce one or two new points. These are critical prerequisities to any successful performance evaluation system.

1. Have the officer sign-off on the evaluation form. This insures that the officer is aware of his specific evaluation and that he knows that he has been evaluated. It does not imply agreement with the evaluation.
2. Do not rely strictly upon any single measure of job performance. Problems of error and difficulties of interpretation exist with all measures and only the use of a variety of indices of performance can tell the whole story of job performance.
3. Feedback can change performance if properly given. It is necessary to give both relative and absolute feedback about job performance. Relative feedback tells the individual officer how well he is doing with regard to other officers in his peer group. Absolute feedback tells him how well he is doing with respect to the ideal officer.

4. Performance evaluation information should be used as one source of data in many personnel decisions. In addition to providing some input regarding the officer's past performance, this helps to serve as a motivating force in promoting good performance in the current job assignment. If an individual knows that merit, as measured by current job performance, will influence such personnel decisions as promotion and salary increases, he will be more motivated to perform well. He will know that good job performance is valued by the department and is properly rewarded.
5. As much time should be spent in counseling good performers as with the poor performers. There is a tendency to ignore good performers and concentrate upon attempts to improve the performance of poorer officers. This may negatively affect the motivation of the good performers. They also need to receive feedback about their performance and to be encouraged to do even a better job.
6. Peer evaluations may be most usefully applied in counseling and feedback sessions. Many departments may be reluctant to use such information in administrative decisions and many officers may not want to act as "judges" of their peers. It is most necessary, then, to assure the officers that their evaluations will be used to help their friends improve their job performance and not to hurt their friends' chances for promotion, etc.

Reference

Landy, F. J., & Farr, J. L. Police Performance Appraisal. Pennsylvania State University, 1975.

END