

PERFORMANCE APPRAISAL IN POLICE DEPARTMENTS

Frank J. Landy

PoliceFoundation

The Police Foundation is an independent, nonprofit organization established by the Ford Foundation in 1970 and dedicated to supporting innovation and improvement in policing. This book is published by the Foundation as a service. The opinions and recommendations are those of the author, and not necessarily those of the Police Foundation.

Copyright © 1977, Police Foundation

Library of Congress Catalog Card No. 77-83722

FOREWORD

This monograph will consider not only the technical aspects of performance appraisal such as the characteristics of forms, the way to treat members, the uses of appraisal information, but also the human problems involved i.e., the reactions of supervisors and subordinates to a system, the motivational properties of a performance appraisal system, and so forth. This can be made concrete by describing how the Dade County Public Safety Department, Dade County, Florida, approached a change in personnel evaluation systems. While no particular experience can be typical, the Dade County experience might help personnel administrators in other law enforcement agencies better formulate an approach to their own unique situation. The last section of the monograph contains suggestions departments may follow for modifying a performance appraisal system or developing a new system.

PREFACE

Since the Police Foundation was established in 1970, its Board of Directors and staff have devoted a large measure of time and resources to issues involving police personnel administration. This concentration on personnel reflects the fact that a significant portion of the efficiency and effectiveness of the police is linked to the selection, training, promotion and supervision of police officers.

So the general subject of police personnel has been a major program area for the Foundation and during the past several years it has sponsored demonstration and research projects in the areas of women in policing, police officer height as it relates to performance, the selection of police chiefs, psychological testing and counseling, and personnel management information systems.

So far, these projects have resulted in several Foundation publications: Policewomen on Patrol (two volumes); Women in Policing: A Manual; Police Chief Selection; Police Officer Height and Selected Aspects of Performance; Police Personnel Administration; and Kansas City Peer Review Panel.

This report marks the publication of a series of monographs on personnel issues. The subjects include performance appraisal in police departments, police selection through assessment centers, and personnel management information systems for the police.

This monograph and others in the series are published in the belief that each can help police leaders and managers in the job of improving the quality and performance of American police personnel.

Patrick V. Murphy
President
Police Foundation

CONTENTS

FOREWORD

PREFACE

I	THE CURRENT STATUS OF PERFORMANCE APPRAISAL	1
II	THE DADE COUNTY EXPERIENCE	15
	REFERENCES	33
	APPENDIXES	35

THE CURRENT STATUS OF PERFORMANCE APPRAISAL

Performance appraisal is an important part of the personnel administration of any police department. However, the current status of performance appraisal systems is discouragingly low both in individual municipal police agencies and in the law enforcement community as a whole. This status may be attributed to perceived inadequacies in some existing performance appraisal systems and to confusion about the place of performance appraisal in a total personnel operation. Such difficulties can cause both supervisors and subordinate patrol officers to object to particular evaluation systems.

First-level supervisors may react to their performance measurement system with either apathy, embarrassment, or qualified support. With the most typical reactions, apathy and detachment, the supervisor refuses to take any personal responsibility for the procedure or the results. When questioned about the system, such a supervisor may say, "The people in personnel want it," or "I have to send these to the captain once a year. I don't know what he does with them." These answers suggest that the supervisor may have had nothing to do with either the construction or the implementation of this system. Often, such a supervisor prefers to disclaim evaluations rather than lose the respect of subordinate officers.

With the second kind of reaction, one of sheer embarrassment, the supervisor cannot be isolated from the personnel procedures of the department in the eyes of subordinates. Consequently, the supervisor apologizes for the procedures, asking subordinates to have faith that things will change for the better. Such a reaction more often occurs in face-to-face encounters between a supervisor and a subordinate in the performance appraisal procedure; for example, when the person being rated must sign the performance appraisal sheet.

The supervisor who supports evaluation procedures is convinced, in the abstract at least, of the value of performance appraisal. Such a supervisor may be familiar with current research and terminology, but may have difficulty applying this knowledge in a particular department. Performance appraisal is not merely a matter of adequate forms and information; it is also a problem of motivation and satisfaction, a problem of supervisor/subordinate relations.

Reactions of subordinate patrol officers can also take several forms. If the officers are relatively new, they accept the performance appraisal system as they do other components of the personnel system. Informal discussions with peers, however, may acquaint new patrol officers with inequities in the system. In a short period of time, the new officers may become aware of discrepancies, illogical decisions, bias, pettiness, and so forth. The acceptance stage then gives way to other reactions--one active, the other passive.

The patrol officer may actively lobby to replace the present system with one ostensibly more equitable. If a patrol officer assumes, however, that the evaluation process could never be fair and unbiased, the officer may even work for its elimination.

The passive reaction replacing acceptance is simply apathy. The patrol officer, like the supervisor, dissociates himself or herself from the procedure, engaging in it for the sake of the personnel department. The officer signs the form each year, because the form states that a signature does not imply agreement with the judgment of the supervisor.

A Short History of Performance Appraisal

Negative responses by supervisors and subordinates to performance evaluation systems may reflect actual inadequacies in those systems. Over the years, performance appraisal has lagged behind other personnel areas in sophistication and development, in spite of its central position in the personnel system. Two other components of the personnel system have received the most attention: selection and training.

Selections procedures have developed greatly since Terman used the Stanford-Binet intelligence test in a study of the selection of police officers and fire fighters in the city of San Jose in 1917. Since that time, the simple spiral omnibus intelligence test has expanded to a test battery often including specific aptitude tests, interest inventories, and personal adjustment tests. In short, the role of testing in the selection of patrol officers has become increasingly complex and important over the last 60 years.

Training academies and programs have become similarly complex. No longer simply learning laws, equipment, and procedures, recruits now receive training in interpersonal relations, conflict resolution, problem solving, and even introductory deductive and inductive logic.

Yet the same departments which use sophisticated selection and training programs may use simple performance appraisal systems, often a series of ambiguous rating scales administered once a year on the anniversary of the employee's date of hire and a check to make sure that the number of absences or late arrivals at roll call has not exceeded some maximum number. In addition, the individual may be required or expected to remain physically fit and proficient with a firearm.

If one thinks about it for a minute, development of personnel procedures seems to be confused. How could selection and training programs have developed before a sophisticated performance appraisal procedure? How can one select a patrol officer for a position without knowing how a successful patrol officer behaves? Selection procedures seem to be based on the assumption that it is cheaper and more logical to use a predictor of performance to select personnel than it is to get a sample of that performance itself. But without good measures of performance, one can never identify good predictors of that performance. Similarly, how can training programs be structured if one does not know how a successful police officer actually behaves? If performance appraisal systems are inadequate, how can a department ever know if its selection and training programs justify the time and expense they incur?

Criteria or measures of performance often exist to validate selection tests. This is exactly the reverse of what should occur.

Department administrators should instead identify criteria and performance components they believe to be necessary for job success; they should try to find or develop tests to predict high scorers and low scorers on those performance measures on the street. The emphasis on tests rather than criteria is natural: A department can take advantage of basic developmental work done by experts in testing. Criteria development, on the other hand, takes a considerable investment of time and energy by a particular department because there are no general criteria instruments available.

Recently, however, police departments have focused on criteria measures and performance appraisal because of court rulings, particularly those relating to fair employment practices. The courts have required that criteria used for the validation of selection tests be demonstrably "job-related." The courts require a clear link between the requirements of the job and measures of performance on that job. Specifically, the courts have suggested that both the selection of tests for a battery and performance measures to be used as criteria for those tests be based on a rigorous job analysis. Furthermore, it has become clear that if performance measures are used to make promotional decisions, those performance measures should be considered selection tests and must be treated as predictors. In addition, increasing education, sophistication, and involvement in departmental procedures of patrol officers and their supervisors have led to questions about the adequacy of personnel procedures related to performance appraisal.

Little research has been devoted to performance measurement in police departments until the last several years. This research (reviewed in Landy, 1974, Landy & Farr, 1975), can be summarized rather easily; the major argument in the research literature has been whether subjective measures of performance should be gathered. Little concern has been shown for the source of performance dimensions, whether subjective or objective.

Nonetheless, there have been at least two notable developments in performance measurement. One occurred in 1960 with the careful development of the Ohio State Highway Patrol forced-choice inventory. It gained wide acceptance in research and development circles but had certain counseling and research limitations, proving most useful in administrative roles. For that reason, most departments remained loyal to the traditional graphic rating scale, although they continued to talk about the forced-choice system.

The second development occurred in 1971 when two projects dealing with performance appraisal in municipal police agencies were begun (Dunnette & Motowildo, 1975; Landy & Farr, 1975). Each project involved both the determination of critical components of patrol officer behavior through job analyses and transformation of these components into performance measurement instruments, quite different from the traditional graphic rating scale. One of these projects (Landy & Farr, 1975) will be discussed later.

In short, until recently, research devoted to performance appraisal as a component in the personnel system of municipal police departments has been almost nonexistent.

The Uses of Performance Measurement

There are three primary uses for performance appraisal information: Measurements of individual work performance may be used for administrative decisions, research purposes, or personnel counseling. Administrative decisions include such subjects as salary adjustments, special duty assignments, layoffs, and promotions. Performance information could also be used in a personnel research program either to validate application tests or to determine whether a training program is effective. Finally, individual supervisors could use performance information to help subordinates achieve personal and organizational goals; that is, the information should be useful for counseling and personal development.

Most departments recognize, at least indirectly, the role of performance information in each of these three processes. When a particular position opens in the department, for example, community relations officer, Officer A seems immediately appropriate for the position because of his "obvious" interpersonal skills. In other words, his past behavior has been noticed and seems to match well the requirements of the new position. In a sense, this process uses performance information to make an administrative decision. Let's say that Officer A actually accepts the job of community relations officer but does not perform adequately in the job. If informed about the substandard performance, the appointing officer may modify decision strategy in the future, no longer quite so confident that officers with good interpersonal skills make good community service officers. In a sense then, the performance information is playing a research role, helping to modify a hiring or placement strategy in the abstract. Finally, because the supervisor would like to help Officer A improve in the community relations position, the

supervisor gathers specific information about Officer A's performance from citizens, other officers, and other supervisors, and sits down with Officer A in an attempt to work out specific performance problems. That is, the supervisor uses performance appraisal for personal counseling and development.

These three processes--administrative decisions, research programs, and personal counseling--go on in every department, regardless of size or circumstance or form of performance appraisal. Formal programs for the development and implementation of performance measurement strategies simply try to improve the efficiency of the three procedures by providing more accurate and reliable data for making the necessary decisions.

The Steps to Performance Appraisal

The first step in a sequence designed to provide information about job performance is job analysis, which identifies the important elements of a particular job. The assumption is that a supervisor must identify the essence of a job before deciding to place a particular individual in that job, determining a research strategy for identifying potential candidates for that job, or counseling individuals already in that job. A job analysis is a description of a job, not of the person in that job. A job analysis helps to determine what standards will later be used to measure the performance of individuals in that job.

There are basically three ways to conduct a job analysis: 1) obtaining information from people who are familiar with the job (the job incumbents themselves or their supervisors); 2) observing the job as it is actually carried out and trying to discern its critical elements; or, 3) actually doing the job and trying to determine its most important elements on the basis of introspection or self-knowledge gained from the experience. The second and third methods are usually impractical; most job analyses are conducted on the basis of talking to people who are familiar with the job. This basic information is then supported with observations and possibly first-hand experience.

The term criterion development takes us one step closer to the actual performance appraisal process. As defined in a dictionary, criterion means some measure of achievement or behavior. Consequently, criterion development is a procedure for identifying or constructing measures of individual performance. As in development of job descriptions and job evaluation, criterion development depends on a job analysis. After the critical components of a particular job are identified, these components must be transformed through criteria development into measures suitable for describing individual achievement. Criterion development is not so clear and distinct a process as job evaluation or job description; it is a process of taking the information provided by the job analysis, which helps supervisors distinguish among jobs, and putting it into a form suitable for making distinctions among people--even those in identical jobs.

Performance appraisal, the actual process of gathering the information about individuals based on important job requirements, is the last step in the sequence. It is a process intended to identify the strengths and weaknesses of individuals engaged in their work roles.

Variations in Performance Measurement

One of the major issues in performance measurement is the nature of the information gathered. Should it be objective or subjective? "Objective" usually means information which is in some way countable or verifiable (e.g., number of arrests, arrest to conviction ratio, or number of tickets written per month). The meaning of the term "subjective" is less clear. Taken at face value, it usually implies that the information is subject to the judgment of the person making the evaluation--the major stumbling block for performance appraisal. While the number of arrests per month is not a matter of judgment on the part of the supervisor, an estimate of the patrol officer's initiative is a judgment. Most critics of subjective or judgmental performance evaluation imply that the judgment made by the supervisor is more related to personal idiosyncracies than to the behavior of the person being rated. Nevertheless, most supervisors would agree that the difference between good and bad patrol officers is most apparent in the intangibles (e.g., such qualities as initiative or attitude), which cannot be counted but must be judged by supervisors or others familiar with the performance.

A third form of information often included in a performance assessment is known as personnel data, because it can be found most often in the personnel folder of a particular officer. It relates more closely to the preconditions or collateral conditions for working (e.g., absence records, tardiness information) rather than the quality of the work itself. But it is closely work-related, inasmuch as an individual who is late or absent is less than 100 percent effective by virtue of that absence.

Undoubtedly no one of these three sources of information can be used exclusively because each has its own shortcomings. The shortcomings of objective data are related to factors, such as shift and beat, that are not under the control of the individual officer, yet influence the numbers which are "counted." The peculiar shortcoming of personnel data is that there do not seem to be enough to go around; only a small proportion (5 percent) of the officers are likely to have any information of a personnel nature in their folders which could conceivably be interpreted as performance-related. This means that the data are "missing" for the rest of the department. Objective data are seldom uncontaminated by situational influences, and personnel data are seldom available for most of the officers in a particular municipality. Therefore, supervisory ratings are often the only measures with sufficient generality to satisfy the administrative, research, and counseling needs of the department. Nevertheless, the peculiar weakness of ratings is their potential for being influenced by characteristics and biases of the supervisor.

Judgmental Data: Approaches and Inadequacies

There are actually several forms of judgmental performance assessment schemes; formal ratings are only one instance. Other familiar judgmental schemes involve checklists and ranking systems.

Check Lists

In this procedure a supervisor chooses from a series of statements describing possible behavior of patrol officers. The value or level of performance described by each of the statements is known beforehand, so it is a simple matter to take the average of the statements checked as an indication of the performance level of the particular officer. Figure 1 shows a list of statements that might be selected to describe a particular officer's performance.

FIGURE 1

A Checklist for Patrol Officer Performance Evaluation

<u>Statements Checked for Officer A</u>	<u>Scale Value^a</u>
1. Goes out of the way to help the public	4.3
2. Is occasionally sarcastic with juveniles	1.9
3. Keeps equipment in top shape	3.2
4. Can be counted on to back up fellow officers	4.2
5. Takes continuing education courses in law enforcement on his own time	4.6
6. Is often late for roll call	<u>1.2</u>
Total of the values	19.4
Average of values	3.47
Value assigned to Officer A =	3.37

^aIn this type of performance appraisal system, the statements have already been assigned values representing the level of performance. In this example, the statements can have values ranging from 1 to 5, with 5 representing outstanding performance and 1 representing very poor performance. Thus, Statement 1 has been previously judged to represent performance at a reasonably high level (4.3) while Statement 2 has been previously judged to represent relatively poor performance (1.9).

Most agencies using a checklist group the items by specific categories, so that a score is obtained for each category. One of the major drawbacks to using a checklist for performance assessment is the time and energy required to develop the items for the checklist and to determine how much of a particular performance category each item represents.

Forced-Choice Scales

A well-known variation of the checklist is the forced-choice format such as the Ohio State Highway Patrol forced-choice inventory. In a typical forced-choice procedure the supervisor examines a list of four statements and picks one of the four as most descriptive of the officer and one as least descriptive. The four items in each group have been carefully chosen for certain properties. Two of the items (one positive and one negative) have been previously shown accurately to identify good and poor performers, respectively. The other two items have been previously shown to vary in favorability (one favorable and one unfavorable) but to have no value in identifying good and poor performers. In terms of its development and scoring, the forced-choice instrument is quite complicated. However, it separates the identification of various levels of performance from a supervisor's attempts to judge an officer favorably or unfavorably, regardless of performance level.

The forced-choice format, like the checklist generally, makes it difficult to identify and place values on many items intended for describing performance. Forced-choice formats exaggerate these problems because of the necessity to determine not only the performance level of each statement and its ability to discriminate among individuals, but also its social desirability value. As a result of these difficulties, individual agencies seldom construct forced-choice scales. Even when constructed, they are generally intended to provide some general procedure for discriminating among individuals on a global level rather than providing individual performance factor scores. This makes forced-choice systems suitable for specific administrative and research purposes, but seldom useful for individual diagnostic counseling or personal development of officers.

Ranking Systems

Ranking systems or employee-comparison methods provide an alternative to checklists or forced-choice systems. While there are many variations of the employee-comparison approach, the two most common are the forced-ranking procedure and the paired-comparison procedure.

In the forced-ranking procedure, the supervisor must place a given percentage of the individuals to be ranked into each of several discrete categories. The supervisor might be instructed to place 11 percent in the top category, 20 percent in the next highest category, 38 percent in a middle category, 20 percent in the next to the lowest category, and 11 percent in the lowest category. Variations in these percentages are not permitted.

In the paired-comparison procedure, each individual is paired with every other individual, and the supervisor must choose the "best" member of each pair. The individual's score consists of the number of times he or she was chosen as the best of a pair. This procedure can become quite tedious if a large number of people must be evaluated. For example, if 12 squad members are evaluated using the pair-comparison procedure, a total of 66 pairs have to be examined and evaluated. This pairing would have to be repeated for each dimension used for evaluation.

Ranking procedures generally have a common flaw: There is no way to guarantee that the best individual in one group of officers would have been the best performer in another group. Appraisal of the level of performance of any one individual is totally dependent on the performance level of the other individuals considered. This makes it almost impossible to make any effective counseling statements which do not include reference to other officers. In addition, the individual department can never determine the absolute effect of a training or motivation program on performance since there is no information about relative or absolute levels of performance.

Performance Ratings

Because of the many problems related to checklist and employee-comparison methods, dependence on graphic rating scales over the years has increased. Performance rating in most departments follows a typical procedure. Most departments have adopted a probationary period of six months to a year for patrol officers, following academy training. During this period, a field training officer or immediate supervisor may rate officers as often as once every two weeks. After the expiration of the probationary period, the frequency of evaluation drops back to the normal department level. Then each officer is rated by an immediate supervisor, usually a sergeant, once a year on the anniversary of hire date. The patrol officer must inspect the rating given by the supervisor and sign the rating form after the inspection, but the signature of the officer does not imply agreement with the rating. The rated officer usually has the right to appeal any performance rating with which he or she disagrees to the next level of command, usually a lieutenant. Once gathered, the performance ratings become part of the officer's permanent file.

Of course, this procedure can vary. Because ratings on individual anniversary dates may use supervisors' time inefficiently, some departments prefer that ratings occur at one time. Some smaller departments have meetings in which several supervisors discuss the performance of all officers and come to a group agreement about the performance level of each officer. The rating form typically looks like the one presented in Figure 2.

Although the number may vary, the rating usually consists of five to ten performance categories, each having five levels, representing a continuum from excellent to poor performance. While the verbal descriptions of each performance level vary from department to department, on the

FIGURE 2
A TYPICAL GRAPHIC RATING SCALE PERSONNEL RATING FORM: CITY OF XXXX

PERFORMANCE CATEGORIES	LEVELS OF PERFORMANCE				
	Always does less than expected	Occasionally does less than expected	Does an average amount of work	Is above average in amount of work completed	Consistently does more than expected
<u>Quantity of Work</u> The degree to which the amount of work matches the standards expected of officers in this department.					
<u>Quality of Work</u> The degree to which the quality of work matches the standard expected of officers in this department	Work is always of poor quality	Occasionally does poor quality work	Work is of average quality	The quality of work is above average	Consistently does high quality work
<u>Work Attitude</u> The degree to which the individual displays a positive attitude about the law enforcement profession	Consistently poor attitude	Occasionally has a poor attitude	Attitude toward work is generally good	Consistently has a positive attitude toward work	Outstanding work attitude
<u>Appearance</u> The degree to which the individual conforms to the guidelines and expectations of the department for dress and general comportment	Presents a poor image	Occasionally must be reminded of	Good appearance	Always presents a good image	Outstanding appearance
<u>Overall</u> The overall contribution the individual makes to meeting the goals of the department	Minimum	Below average	Average	Above average	Maximum

average these descriptions or "anchors" consist of short evaluative phrases, such as "above average in work quality," or "quantity of work occasionally falls below standard."

Guion (1965) suggested that rating scales, such as those presented in Figure 2, can differ in at least three important respects: 1) The nature of the definitions for each category; 2) the nature of the statements used to describe each level of each category; and 3) the nature of the response made by the supervisor.

In Figure 2, the definitions for the categories lack clarity. For example, the statement about "quality of work" does not give specific details. In addition, statements describing the levels of each category do not give the supervisor enough information. For example, what does "quality of work is above average" mean? This statement does not specify how much "above average"; for that matter, it allows the supervisor to define "average"; finally, the meaning of "quality" itself may be unclear. The clarity of rating scales may occur only in the procedure: The supervisor must check one of five boxes for each category and has no alternatives to this procedure.

Consider the differences between rating scales for "quality of work" as shown in Figure 2 and as shown in Figure 3.

A supervisor, using the "quality of work" scale in Figure 3 rather than the corresponding scale in Figure 2 probably assesses an officer more accurately. This difference in "accuracy" pinpoints the disadvantages of a rating procedure for performance assessment. If the scale used is technically inadequate, the numbers obtained from that scale will be useless. Not all performance assessment problems will be solved with a technically adequate rating scale, but with a soundly constructed scale a department can gather accurate performance information.

If scales are not well designed, one may encounter at least one of the following problems: Rating "halo," rating "leniency," and "central tendency." Each of these problems will be described more fully below.

Halo. The term "halo" means that a general response on the part of a supervisor to a particular officer affects all judgments about that officer. For example, a supervisor may rate an officer on five different scales, but assign all ratings on those five scales at the same level of performance, i.e., 4 4 4 4 4 or 3 3 3 3 3. Any scale which allows a supervisor to depend on general impressions may lead to halo errors. Clear definitions of categories and explicit anchors for levels of performance help prevent halo errors.

Central Tendency. Central tendency errors occur when ratings for all officers in a department artificially center around one performance level. This point is not necessarily the arithmetic average of the scale (e.g., the middle category), but the "subjective" center of the scale which

FIGURE 3

A GRAPHIC RATING SCALE WITH AN ADEQUATE CATEGORY DEFINITION AND ANCHORS

PERFORMANCE CATEGORY	LEVELS OF PERFORMANCE
<p>Quality of Work Consider the care with which the individual performs required duties. Is it often necessary to have someone else check this officer's reports? Does the officer fail to follow guidelines or arrest procedures, making prosecution difficult?</p>	<p>Frequent omissions and inaccuracies in reports; requires constant and close supervision.</p> <p>Occasional errors in work but is capable of correcting them when they are pointed out.</p> <p>Occasionally makes minor errors but they are usually due to recent changes in regulations or procedures.</p> <p>Seldom makes errors in the application of the law or department procedure. Requires little supervision; often makes work to eliminate errors.</p> <p>Consistently error-free work. Always checks work twice. Can be counted on for high-quality work without supervision.</p>

typically is "4" on a five point scale. In other words, everyone in a department is "above average" on every dimension. In many departments, procedures encourage this error. If an officer receives a below average or outstanding rating (1, 2, or 5) a paragraph or more must justify the score. This leaves only the levels of performance represented by the numbers "3" and "4". If a supervisor does not see much difference between these levels and generally favors an officer's performance, a supervisor will probably assign "4" as the performance level. A rating of "4" also will not disrupt comfortable relations between supervisor and subordinate, a situation which exposes the essential problem of trying to make subjective supervisory ratings objective. Though personal or procedural reasons may encourage a supervisor to choose the subjective middle of the scale, vague statements describing the levels of performance permit the central tendency error to occur. Such statements are general enough to allow the level of performance rating to depend on a supervisor's general impression of the department rather than a particular officer's behavior.

Leniency. Patrol officers may describe certain supervisors as particularly strict or particularly lenient in the assignment of ratings, giving low or high ratings respectively. In other words, if one could accurately measure any behavior in a group of patrol officers, the ratings of a strict supervisor would be consistently lower than the actual performance of the patrol officers; the ratings of a lenient supervisor would be consistently higher. Again, a vague scale permits this kind of error to occur. If the levels of performance are not well described, the supervisors must produce their own standards for rating an individual. Generally, strictness is the culprit when patrol officers complain about the "personality factor" in ratings.

These three types of errors usually occur in combination with one another, so that ratings in most departments vary little. Variance usually results from the characteristics of a supervisor rather than from those of subordinates. In a typical department with a series of five-point rating scales, probably the rating on any dimension for any patrol officer chosen at random will be "4". When such a situation occurs, both supervisors and subordinates recognize the inadequacy of a performance appraisal system. The final section of this monograph presents some guidelines and suggestions for developing satisfactory procedures for adequate performance appraisal.

The Future of Performance Appraisal

A performance appraisal system must not only satisfy a department's personnel, but must also bear the weight of supporting decisions which may be scrutinized in the course of court cases of alleged discrimination. Further, the identification of performance quality is necessary for motivational programs of personnel and personal development. This last use highlights the future of performance appraisal systems.

Patrol officers, becoming more aware of the complexities of their roles, realize that they perform in overlapping roles requiring different skills at different times. The maintenance of these skills and the development of new ones require feedback of the most sophisticated sort rather than global ratings of effectiveness. Almost all of the current theories of work motivation depend on an accurate and discriminating assessment of various aspects of work performance. For example, a Management By Objectives (MBO) Program depends on the identification of individual strengths and weaknesses; it also depends on the identification of changes in these levels over time. A well constructed performance measurement system can provide such data. An improved performance appraisal system may fulfill the traditional goals of research and administrative decision-making, and an improved system can make a unique contribution in the form of providing the information necessary for the implementation of motivational programs.

THE DADE COUNTY EXPERIENCE

Dissatisfaction with one personnel evaluation system may lead a department to replace it with another. When the Dade County Public Safety Department developed a new appraisal scheme, several problems occurred before a satisfactory system evolved. Familiarity with procedures followed in Dade County may help other departments anticipate and deal with these problems.

The Dade County Public Safety Department

Dade County is an area of more than 2,400 square miles in Florida, encompassing 27 municipal governments. The area's population is constantly increasing, and currently includes more than one and a half million full-time residents. Miami itself has maintained one of the fastest growth rates in the country; its population has increased 35 percent since 1960.

The Dade County Public Safety Department under Director E. Wilson Purdy has a department of 1,100 police officers, 258 sergeants, 78 lieutenants, 23 captains, and over 600 support and clerical personnel. The physical properties of Dade County range from farm to city; from grove and farmland to high-rise hotels, motels, and apartments; from marinas to race tracks; from country trails to interconnecting expressways. A police officer in Dade County must be trained to handle the interests of citizens in these diversified areas and those of more than seven million annual visitors.

New officers in the department receive 20 weeks of basic training. In-service training continues in such areas as human relations, management, organized crime, and so forth. Public safety officers are encouraged to continue their formal education. As a result, the department currently has one of the highest educational levels of any major police force in the nation.

Dade County operates under a merit system with a central personnel office in charge of county-wide testing, classification action, and other personnel functions. A personnel bureau within the Dade County Public Safety Department administers personnel records, leave and attendance records, payroll, and assists employees in all general personnel matters while acting as a liaison to the central personnel office. This bureau also screens and recommends for employment those applicants who have passed the initial written test.

The personnel bureau, a subdivision within the department's administrative division, reports directly to the administration division chief. One of its primary objectives has been to conduct personnel-related research in the areas of performance appraisal, biographical data,

validation of the oral interview system and development of a management information system. The bureau consists of 6 administrative and 14 clerical personnel, though three outside consultants have assisted in the work in varying degrees. The Public Safety Department, although completely responsible within itself for all programs developed to bring about personnel-related changes, must, however, clear all such changes with the personnel director of Dade County who is responsible for assuring the appropriateness of the individual department's actions. To date, the central personnel department has been most cooperative and willing to accept innovative personnel procedures. Currently, the staff is assisting the Public Safety Department in creating performance evaluation forms for police sergeants and police lieutenants.

The Old Rating Form

At the beginning of the project, the Public Safety Department was using a typical performance rating system. Figure 4 shows the form and the instructions given to supervisors. The behavior categories are not well defined, and the levels of performance are not well described. Although the marking system is rather straightforward, officers often wanted to use more than five response categories and attempted to alter the scoring format to give variations on the five-point scoring format. Such variations were discouraged. Ratings, gathered once a year on the anniversary date of hire, became part of the officer's permanent personnel folder. However, the ratings seldom were used for administrative decision making; they were not useful as research tools and were rarely effective for personal counseling and development. In short, they were religiously administered, filled out, and filed--and just as religiously ignored.

The inadequacy of the performance appraisals is evident when the actual ratings themselves are examined. The average intercorrelation among the five dimensions is about .95. This means that, for all practical purposes, a rating on any one of the five dimensions is as good as ratings on all five of the dimensions; that is, the five ratings do not represent five distinct areas of performance. There seem, instead, to be four repetitions of one dimension.

Furthermore, leniency and central tendency effects combine so that almost every rating is a "4". This means that if one checks the ratings of any officer for any performance category, the number will most likely be a "4".

This lack of rating variation predictably affects three possible uses of performance measurement. It means, first, that the ratings are useless for research purposes. For example, attempts to validate the Dade County entry level tests and selection procedures, such as the oral interview, consistently yielded nonsignificant correlations. In a system in which variations in test performance should predict variations in actual work performance, variation in both the predictor and criteria is essential. With no variation in criteria measures (ratings), nothing can be predicted and the correlations must be zero.

FIGURE 4

**METROPOLITAN DADE COUNTY
EMPLOYEE PERFORMANCE REPORT
WORK SHEET**

Name (Last) (First) (Initial)			Period Covered From To	
Civil Service Title		Civil Service Status	If Prob, Date Ends	
Department		Division	Unit	

<p style="text-align: center;">CHECK ITEMS</p> <div style="display: flex; justify-content: space-between; font-size: small;"> <input type="checkbox"/> Strong <input type="checkbox"/> Weak </div> <div style="display: flex; justify-content: space-between; font-size: small;"> <input checked="" type="checkbox"/> Satisfactory <input type="checkbox"/> Not applicable </div>	INDICATE FACTOR RATING BY "X"			
	UNSATISFACTORY	NEEDS ATTENTION	SATISFACTORY	OUTSTANDING
<p>1. QUANTITY OF WORK</p> <p><input type="checkbox"/> Amount of work performed</p> <p><input type="checkbox"/> Completion of work on schedule</p>	Seldom produces enough work or meets deadlines.	Does not always complete an acceptable amount of work.	Consistently completes an acceptable amount of work.	Amount of work produced is consistently outstanding.
<p>2. QUALITY OF WORK</p> <p><input type="checkbox"/> Accuracy</p> <p><input type="checkbox"/> Effectiveness</p> <p><input type="checkbox"/> Compliance with instructions</p> <p><input type="checkbox"/> Use of tools & equipment</p> <p><input type="checkbox"/> Neatness of work product</p> <p><input type="checkbox"/> Reports & correspondence</p> <p><input type="checkbox"/> Thoroughness</p>	Too poor to retain in job without improvement.	Quality below acceptable standards.	Performs assigned duties in a satisfactory manner.	Performs all duties in an outstanding manner. Exceptional accuracy, skill or effectiveness.
<p>3. WORK HABITS</p> <p><input type="checkbox"/> Attendance</p> <p><input type="checkbox"/> Observance of working hours</p> <p><input type="checkbox"/> Observance of rules</p> <p><input type="checkbox"/> Safety practices</p> <p><input type="checkbox"/> Personal Appearance</p>	Too poor to retain in job without improvement.	Work habits need improvement.	Work habits satisfactory.	Exceptional work habits. Always observes rules and safe practices.
<p>4. PERSONAL RELATIONS</p> <p><input type="checkbox"/> With fellow employees and supervisors</p> <p><input type="checkbox"/> With public</p>	Too poor to retain in job without improvement.	Personal relations need improvement.	Maintains satisfactory work relations with others.	Exceptionally co-operative with public, co-workers and supervisors.
<p>5. SUPERVISORY ABILITY</p> <p><input type="checkbox"/> Planning & assigning</p> <p><input type="checkbox"/> Training & instructing</p> <p><input type="checkbox"/> Disciplinary control</p> <p><input type="checkbox"/> Evaluating performance</p> <p><input type="checkbox"/> Delegating</p> <p><input type="checkbox"/> Making decisions</p> <p><input type="checkbox"/> Fairness & impartiality</p> <p><input type="checkbox"/> Unit morale</p>	Poor supervisory ability. Work of unit frequently unsatisfactory.	Supervisory ability inadequate in some respects. Works results of unit below par at times.	Obtains good results from subordinates. Controls unit efficiently.	Outstanding ability to get maximum from unit and available resources.

FOR SUPERVISORS ONLY)

RATER'S COMMENTS: (attach additional sheets if needed)

RATER'S RECOMMENDATION (for employees under consideration for a merit raise or permanent status)

This is to certify that the overall performance of the subject employee ☐ is ☐ is not satisfactory

The employee ☐ is ☐ is not recommended for
☐ a merit raise ☐ permanent status.

This report is based on my observation and knowledge. It represents my best judgment of the employee's performance.

RATER _____ Date _____

I have reviewed this report. It represents the facts to the best of my knowledge. I concur in the recommendation, if any, as to merit raise or permanent status.

REVIEWER _____ Date _____

In signing this report I do not necessarily agree with the conclusions of the rater. I understand that I may write my comments on the reverse side. I have received a copy of this report.

EMPLOYEE'S
SIGNATURE _____ Date _____

Second, the lack of rating variation affects administrative decisions. Because the ratings do not distinguish one individual from another and because all of the ratings point to above-average performance for all individuals, the ratings are worthless for making duty assignments, salary decisions, promotional decisions, and so forth. In short, administrative decisions in the Dade County department had to be made just as if the ratings did not exist.

The third use of performance ratings, personal counseling and development, did not benefit from the rating system because the ratings could not distinguish strengths and weaknesses in individuals. The ratings gave exactly the same information--that a particular officer was doing "just fine" in all behavior categories. Also, because the line patrol officers could see individual differences among members of their own squads, identical ratings affected motivation negatively; better officers were not given adequate recognition for their performance. Furthermore, the few slight rating variations among officers were often attributed to a supervisor's characteristics rather than an officer's behavior. Distrust in the significance of any ratings made the ratings ineffective in personnel counseling.

Besides these problems, the rating form itself lacked face validity because it was used for all county employees regardless of department; it was very general and not tied to the responsibilities of any one job. The police officers resented being evaluated on the same forms as clerks, sanitation workers, cashiers, and so forth. They felt the unique job of a police officer required a unique measurement system, not one intended for all county employees.

The Impetus for Research

It might be useful for other departments to identify the different forces acting on the Dade County Public Safety Department at the time which contributed to the research and development of a new performance appraisal system. Besides a growing dissatisfaction and frustration with the traditional evaluation scheme, the department needed and was developing an aggressive program of overall personnel administration.

Professionalism. The young and heterogeneous Dade County police force had a high educational level. The patrol officers and their immediate supervisors were increasingly aware both of the complex role of the police officer and the advances in personnel technology in other departments. The traditional civil service form gave insufficient information to the officers for personal development for their jobs.

Selection. Similar to most municipal law enforcement systems at the time, the Dade County Department used a rather complicated selection battery (including, among other components, a civil service exam, a physical exam, an interview, and a background check). However, the validity information was discouraging because there seemed to be little or no

relationship between the selection tests and performance ratings. Validation difficulties encouraged a change in personnel policies because of the number of minority members (primarily females, black, and Spanish-surnamed applicants) seeking positions in the department. Fair employment practices are difficult in a prediction system without validity; thus, practical, legal, and moral considerations demanded the construction of a selection system to ensure equal employment opportunities for all applicants. Clearly, constructing such a system required replacing the performance measurement system then in use.

At the time, plans had been made to develop a selection system with two new procedures: Using biographical information in an attempt to predict future performance, and administering an oral interview for applicants. Although nearly 200 questions were answered by each of 300 patrol officers, biographical data subsequently failed to predict patrol officer performance with a high degree of accuracy. In his bio-data project report, available from the Dade County Public Safety Department, Dr. William Buel stated that alternative validation procedures may have produced more encouraging results. Because both the biographical data and interview procedures were based on careful job analyses, the only component of the prediction system still missing was the performance appraisal component.

Management Information System. The size and complexity of the Dade County Public Safety Department required a sophisticated police personnel management information system. Necessary information in such a system would be individual performance measures, not only the traditional objective performance information (e.g., number of arrests or citations issued), but also judgmental information in the form of ratings. Such judgmental information was deemed essential for the complete description of an individual and necessary for the effective use of a management information system.

The completed police personnel management information system included almost 200 items of information for each of almost 1,300 sworn personnel. Subsequently, it has proven extremely valuable in administrative decision-making. The system and its development have been fully described by Dr. Wayne Cascio in his monograph Police Personnel Management Information Systems: The Dallas and Dade County Experiences (Washington, D.C.: Police Foundation, 1977).

The Research Instruments: Background

In 1972, the Law Enforcement Assistance Administration (LEAA) of the United States Department of Justice funded a comprehensive study of the status and future of performance appraisal in law enforcement agencies. In particular, LEAA was interested in the development of a judgmental rating system which would satisfy all the uses of performance measurements, i.e., test validation, administrative decision making, and

counseling. The study, conducted over a period of four years at The Pennsylvania State University by this author and a colleague, Dr. James L. Farr (Landy and Farr, 1975), is important for understanding the background of the Dade County developmental research.

It had been concluded early in the LEAA research project that most existing rating systems in police departments were inadequate. The goal of the project was to develop rating scales for the job of police officer that would be accepted by individual supervisors and officers, that could be modified to fit the needs of any department, and that were technically sound. Behaviorally anchored rating scales were chosen as those most likely to serve the needs outlined above. A quick glance at Figure 5 shows how these scales differ from the more traditional scales. This particular type of scale has many positive properties: The dimension definition is adequate, the anchors are descriptive of levels of the performance dimension, and, since the supervisors are requested to mark only at full or half-point intervals, the actual rating is made with a minimum of confusion.

In spite of their rather unusual appearance, behaviorally anchored scales differ from traditional scales more in development. They are based on the identification of critical incidents in the performance of people in a particular job, in this case, patrol officers. These incidents should clearly differentiate the good performer from the poor performer. The identification of these critical incidents and the cataloging of the behavior which these incidents represent is the operation of job analysis, as described earlier. The behaviorally anchored rating scale procedure further orders incidents describing a particular performance category according to the proportion of the particular category they represent.

The actual steps in developing a behaviorally anchored rating scale can be seen rather clearly in Figure 6. While Dade County was not directly involved in the early phases of scale development, a detailed examination of the entire process is necessary to understand the point at which Dade County became involved in the project.

The first step in the process is a conference at which all of the important dimensions of performance are identified. As part of the larger LEAA project, such a conference in Chicago was attended by eight supervisory officers from eight different police departments. Over a period of six hours, these participants developed labels and definitions for the major categories of patrol officer performance. A second conference was held with eight different patrol supervisors. The results of the two conferences were compared and those categories and definitions appearing in both conferences were retained for further development.

Next, another group of patrol supervisors was asked to write behavioral examples of each of the performance categories identified and defined by the earlier conference groups. Once again the participants

FIGURE 5

EXAMPLE OF A BEHAVIORALLY ANCHORED RATING SCALE

Use of equipment: Knowledge of and skill in the use of firearms and other special equipment.

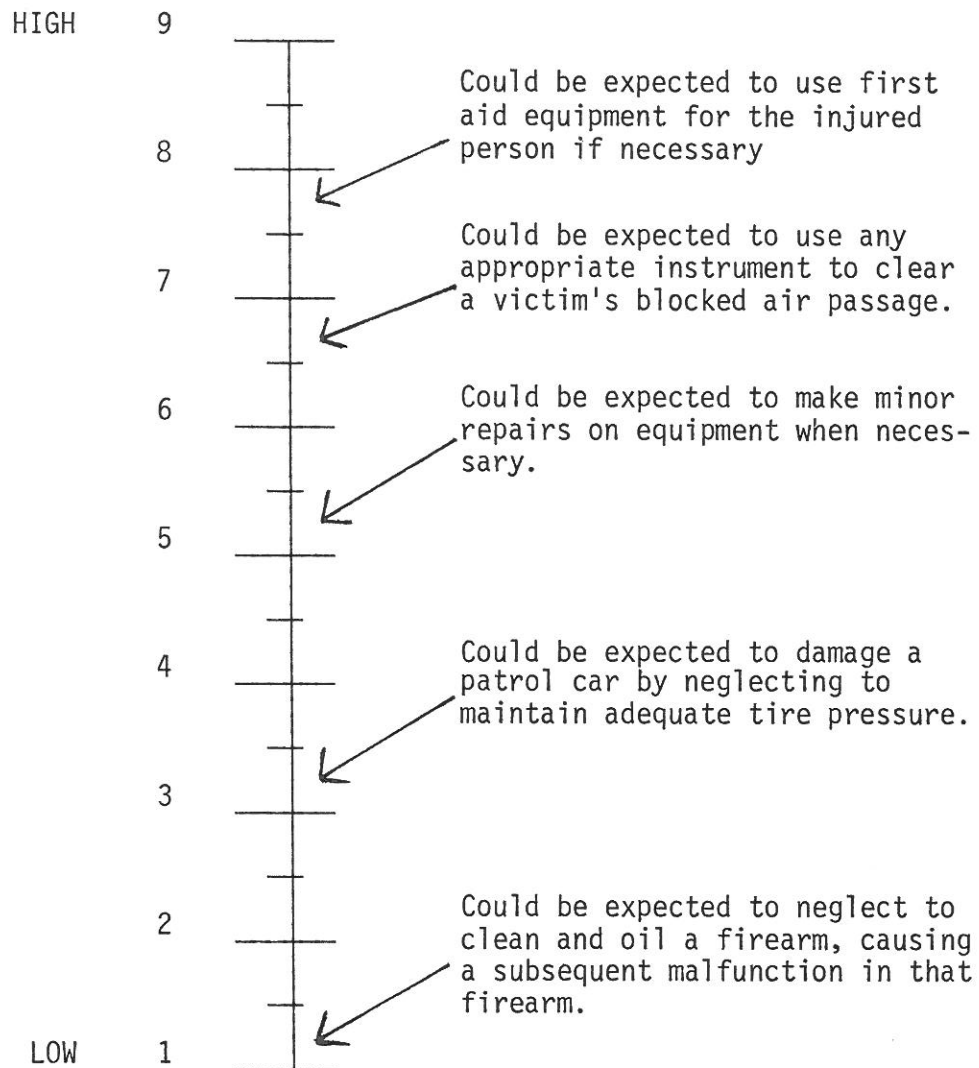
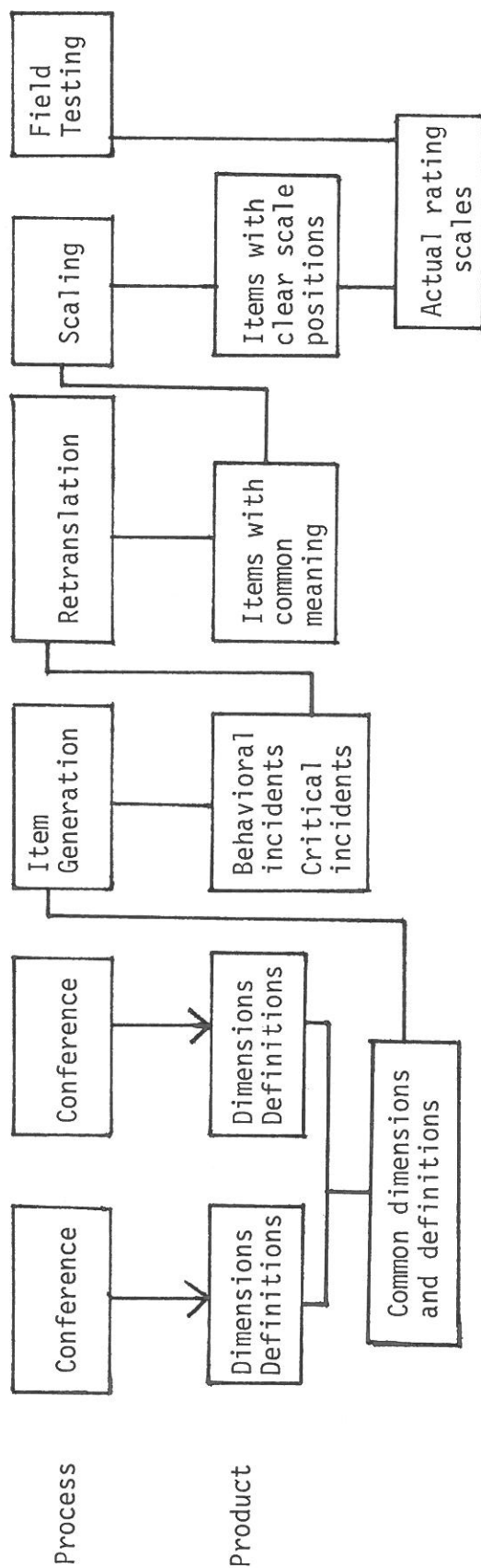


FIGURE 6

PROCEDURE FOR DEVELOPMENT OF BEHAVIORALLY ANCHORED SCALES



represented many different agencies. This step was carried out both by mail and in person.

The third step was to test the behavioral examples with another group of supervisors asked to identify the category for each of the examples. This phase is often called "re-translation" because of its similarity to a common procedure for checking the adequacy of a translation of material into a foreign language. After someone translates the material into a foreign language, another individual translates it back into the original language to see if the message remains unchanged. If the two original language versions of the material are identical, one may conclude that the translation was a good one. In the development of behaviorally anchored rating scales, the procedure is similar. A group of experts is asked to place behavioral descriptions into the categories for which they were originally written; in effect, this checks whether particular items have some generally accepted meaning. The re-translation procedure identifies ambiguous or meaningless items which are then eliminated from the pool of behavioral descriptions.

The next step was to take behavioral descriptions written for the specific categories which had survived the re-translation and to order them in terms of importance (i.e., to assign scale values to them). Another group of patrol supervisors evaluated the descriptions listed under a particular behavior heading and assigned a number (e.g., from a low of 1 to a high of 9) to indicate how much of a particular behavior was represented by the specific description. Items on which experts could not agree were eliminated. The remaining items formed the basis for the scales when arranged in a manner similar to that shown in Figure 5. Supervisors then tested the rating scales in performance appraisals of subordinates. These ratings were examined for the degree of halo, central tendency, reliability, and leniency connected with the use of the scales.

The procedure for developing the behaviorally anchored scales has several important advantages. The first is the demanding nature of the step-by-step procedure; each step is a check on the earlier one. The second advantage, both technical and motivational, is the heavy involvement of those who will actually use the scale. Technically, people most familiar with the job can best identify specific behaviors for each performance category. Also, when the supervisors help develop a procedure, they become motivated to take steps to insure its success, if and when it is put into operation.

The Dade County Involvement

The Police Foundation and Dade County Public Safety personnel staff, both aware of the existence of the LEAA project, saw advantages in using the behaviorally anchored scales. Because the LEAA project had been underway for more than a year prior to the Dade County bio-data and management information system projects, considerable savings in time and effort

for the department might be possible. In addition, the Dade County department could compare itself with the large municipal law enforcement departments participating in the LEAA research. Consequently, the LEAA staff was invited to explore with Dade County the possibilities of introducing the scales into the police department.

Late in 1973, the personnel department of the Dade County Public Safety Department reviewed the Pennsylvania State University's work on the development of the scales for performance appraisal. The conferences had been held, the behavior categories defined, and the critical incidents had been written, re-translated, and scaled. In short, a job analysis had been completed and the rating system was ready for construction.

At the time, the Dade County representatives expressed reasonable concern that the already developed information might not be in their situation because the job analysis and critical incident production occurred in locations other than Dade County. Subsequent development procedures checked that the Dade County patrol supervisors defined the job of patrol officer in the same way as did the supervisors from agencies involved in the earlier conferences. To check the applicability of items and definitions, in March 1974 a sample of 45 Dade County patrol supervisors scaled 80 descriptions of critical incidents. The mean scale value for each of the 80 items averaged across the 45 sergeants was calculated. These 80 values, correlated with similar values obtained from three other agencies of similar size, but in Northeast and Midwest locations, yielded a 4 x 4 correlation matrix (each agency with each other agency). The average correlation of the Dade County scale values with those from the other agencies was .95, strongly suggesting that the initial work done in other locations was applicable in Dade County.

In addition to the data provided through the rescaling, the Pennsylvania State University project staff asked the supervisors if the most important dimensions of patrol officer performance had been identified. Although all the supervisors believed that the proper dimensions had been identified, many of them disagreed with some of the behavioral descriptions, particularly some of the negative ones. The supervisors felt that anyone who acted negatively did not belong on any police force anywhere. The very presence of some of the descriptions seemed to offend them. Attempts were made to assure the officers that a complete scale must include all aspects of performance--not only the exceptionally good aspects, but also the exceptionally bad aspects. For the supervisor to understand the range of levels in each behavior category, the project staff considered it necessary to retain both negative and positive descriptions. The staff feared that a scale consisting of only positive descriptions would limit the technical adequacy of the eventual ratings. In addition, because these descriptions had come directly from patrol officers in other departments, the staff had some confidence that these

negative items represented real behavior that might face the Dade County department in the future.

The next phase in scale development tested several different formats for the rating instrument, a behaviorally anchored scale and a mixed standard scale. In the latter checklist format, the supervisor considers several statements and decides if the person being rated is better than, equal to or worse than the performance described by each item; 24 statements (one high, one average, and one low behavior description from each of eight previously defined performance categories) were presented in such a random check. The supervisors were asked to rate several of their subordinates using both types of rating scales, the behaviorally anchored scales and the mixed standard scales.

Thus, available data enabled a direct comparison of the adequacy of the two procedures. Of particular concern was the relative reliability of each procedure, whether two different supervisors could use a rating procedure to describe the behavior of a subordinate and agree on that description. Forty-two patrol officers were identified who could be evaluated by pairs of supervisors using both formats. The calculation of interrater reliabilities for the two different procedures yielded strong support for the superiority of the behaviorally anchored format. The interrater reliabilities for the mixed-standard format were very low, averaging about .20. For the behaviorally anchored format, these values were more acceptable, averaging approximately .60. The scales were subsequently put together in the behaviorally anchored format for a full field test.

During the month of August the Dade County Department ran supervisory training sessions at which the new behaviorally anchored rating scales were introduced. Each supervisor was asked both to rate one or more subordinates and to comment on the scales and their possible use in the department. A total of 389 patrol officers were rated with the scales during the month.

The data analysis indicated that the resulting scales successfully eliminated traditional halo, leniency, and central tendency errors, particularly when compared to the previously used scales. Nevertheless, supervisors continued to complain about particular aspects of the scales, feeling, for example, that anchors on the scales were out of place (i.e., that a particular behavior description should have a higher or lower position on a scale). Such uneasiness is a usual response to changing organizational practices.

Consequently, in the fall of 1974, representatives from the Dade County department joined with representatives from other departments to present a critical analysis of the problems with the scale format. Though the scales were demonstrably sound, they would not be accepted, as they stood, by the supervisors in Dade County. The critical-analysis conference proposed to identify format changes that would satisfy the

objections of the supervisors without decreasing the technical adequacy of the scales.

Discussions in that meeting identified three problems. The first problem was the objectionable nature of some of the negative behavior descriptions (for example, behavior involving lying or cowardice). It was agreed that other less offensive items with similar scale values could be found.

The second problem had to do with the specificity of the behavior descriptions. In Figure 5, the arrow is intended to tell the supervisor the scaled value of the behavior descriptions. However, since each item was placed on the scale at the average of the judgments made in the scaling phase, the absolute placement of the arrow was misleading. It was agreed to remove the arrows and to note its effect on the technical characteristics and the acceptability of the scales to the supervisors.

The final problem had to do with the phrase "could be expected to" which preceded each of the behavioral descriptions. The original reason for the phrase was to indicate that the behavioral descriptions represented examples of the performance categories rather than absolute definitions of levels of the dimension.

To the raters, however, the phrase implied that they should predict the future, which they were both unwilling and unable to do. They rightly understood performance appraisal as a measure of past behavior, not an anticipation of future behavior. It was agreed to delete the phrase "could be expected to" and to assess the effect of the deletion on the acceptability and technical adequacy of the scales.

Results

In February 1975, 87 supervisors used the newly constructed scales with the changed format to rate subordinates. There were few complaints about item positions now that the arrows had been deleted. There were no complaints about predictive implications of the descriptions now that the phrase "could be expected to" had been deleted. And finally, since the objectionable negative items had been replaced with less objectionable ones at a similar level, there were no complaints about specific items. In short, most of the objections of the users of the scales had been eliminated by the changes. More importantly, the technical characteristics of the scales remained unchanged. Halo errors, leniency errors, and central tendency errors remained few. In addition, reliability was not reduced by the changes. The changed format was ultimately evaluated in over 25 cities and these data confirmed that the changes had made the scales acceptable without altering their technical adequacy. The final version of the scales appears in Appendix A along with instructions for their use.

In a conference in March 1975, all of the accumulated data concerning the development of the scales in Dade County were presented to the command staff which decided to put the scales in operational use in the department immediately. In October 1975, all police supervisors began training in the use of the new scales, whose purpose and method of development were stressed. The new scales are currently used by the department for more than 800 patrol officers. In February 1976 work was begun to develop similar scales for police sergeant and police lieutenant. An initial rating form for police detectives has also been developed but is not yet being used. On June 1, 1976, a new administrative order was released outlining the current evaluation system within the Dade County Public Safety Department. (See Appendix B).

After the new procedures had been followed for several months, a sample of 300 old ratings and 300 new ratings from the personnel folders of patrol officers were examined to determine if the new scales discriminated among the performances of officers better than the old scales. The old scales showed no variation across performance categories; approximately 96 percent of the ratings were "satisfactory." The new forms showed much wider variation; the standard deviations for the ratings were approximately 1.65 on all dimensions (computed on a nine point rating scale), indicating a full use of the scale values. These results suggested that the supervisors could better describe the strengths and weaknesses of the patrol officers with the new scales than with the old scales.

One immediate research use for the scales was an attempt to validate the oral board interview used for the selection of applicants to the academy. Validation attempts using the old rating form had failed, but with the new ratings as criteria, oral board scores were shown to be valid predictors of future performance.²

The Development Process

Let us assume that discouragement with or analysis of an existing system leads to the conclusion that a change of systems is necessary. What are the steps to be taken in the development and implementation of a new system? The development activities can be put into two distinct categories: 1) technical, and 2) interpersonal or motivational.

Motivation and the Involvement of Line Personnel

The importance of involving the supervisors and the patrol officers in the development of a performance appraisal system at the earliest stages and maintaining that involvement cannot be overemphasized. Almost

²F.J. Landy, "The Validity of the Interviewed in Police Officer Selection," Journal of Applied Psychology 61, April 1976, 193-98.

all manufactured products are presently designed with the needs of the consumer in mind; panels of consumers are often formed at each development step to make sure that a new product has properties either useful or pleasing to the consumer.

Supervisors and patrol officers, like personnel administrators and higher level line officers, can be considered consumers for a performance appraisal system. Each category of user has needs and preferences for a performance appraisal system that an effective evaluation procedure should attempt to satisfy.

Furthermore, involving in the development process those who will eventually use a system can contribute to the success of that system. If a person invests time and energy in some process, one assumes that person will want success for that process. The greater the number of people involved, the greater the degree of commitment; conversely, the fewer people involved, the less concern for success of the process. If an individual department decides to adopt a new system without the active involvement of supervisors and line patrol officers, little or no active cooperation should be expected and active resistance might even occur. Naturally, other serious problems can occur when an entire department participates in the developmental process at one time--nothing might be accomplished. Nevertheless, a system in which supervisors and patrol officers elect or appoint representatives to participate is not always a productive solution. Often those not elected or appointed feel they cannot influence the outcome of the process. Supervisors and patrol officers must feel that they have the opportunity to become involved in the development process at any stage, if they so desire. Perhaps volunteer development sessions could be held frequently for consulting the department or perhaps individuals wishing to participate could be assigned various tasks in the process, or at least be assigned to consulting groups at various stages in the development of the appraisal system.

Efficient participation of department members requires both familiarity with technical information and continued feedback about the development process. One or more training sessions can familiarize supervisors and officers with the technical concepts and vocabulary involved in the development process. This background can be given in one training day or, if necessary, over several roll-call sessions, though the single training day is more efficient by far. During this training, the officers learn about the role of the performance appraisal process in a personnel system, common forms of rating, rating errors and the specific aims of the department in constructing a new procedure; only then can officers and supervisors help generate the specific components of a new system.

Once officers and supervisors join in the process, continued feedback about the development efforts is crucial if the commitment generated by participation is to continue until the new evaluation system becomes a departmental procedure. If an individual is involved in early phases of the project, that individual must be kept informed of what happens after

that involvement. The entire development procedure should be open, with no steps taken or decisions made that could not be announced to the entire department.

These requirements for the initiation and maintenance of involvement are demanding. Nevertheless, they must be fulfilled if the new system is to be effective.

Supervisors and subordinates should not just assist the development of a performance appraisal system; they should also formulate guidelines for the evaluation of the new system. Such guidelines must be determined before the new system is constructed. Unless all concerned parties can agree on what constitutes a better system before any new system exists, problems concerning the definition of success will inevitably result when the development effort must be evaluated.

There are several aspects in defining the success of a performance appraisal system. A technically successful new system should have better reliability, less leniency, less halo, and less central tendency embedded in it than in the old system. Furthermore, the successful new system should produce information specific enough to validate the selection procedures used by the department.

Another aspect of success concerns the reactions of the users. A clearly specified procedure for systematically obtaining reactions to the new system should involve a comparison of the old system with the new system. In addition, the time schedule for such an evaluation should be fixed, even to the point of picking specific dates for gathering this information because such an agreement can insure time for an adequate evaluation. A firm agreement makes concerted efforts to terminate the development before its concluding evaluation unlikely.

With good involvement of personnel from the beginning, obtaining agreement about both the definition of success and the timing of the evaluation should not be difficult. As a matter of fact, such a structure may significantly help to provide the real and immediate goals necessary for maintaining interest and motivation over an extended period of time.

Technical Considerations in the Development Process

There are many ways in which the development of a performance appraisal system might be structured. A description of these procedures could be either general enough to accommodate all the various forms of development or very specific, presenting one concrete way of going about it. The following guidelines are a compromise between these two positions. First, a necessary step is phrased broadly enough to cover variations of that particular phase of development. Immediately following is a specific example of how that step might be made concrete. This is not meant to imply that the specific example is the only way to meet the demands of the steps; it is only one way. Given four steps and the concrete examples, it should be possible to find other solutions suitable for particular agencies.

1. Job Analysis: Determine the aspect of performance to be measured

This goal might be best achieved by identifying an experienced, flexible, and interested group of supervisory and patrol officers and asking them to identify all of the possible categories of performance that define the job of patrol officer. (In Figure 6, this appears as the "Conference" block on the left side of the diagram.) First, ask the participants to label the categories, e.g., initiative, judgment, knowledge of equipment. A group of seven or eight individuals should be able to produce more than 20 such labels. Let them continue to produce those labels until no more are suggested. Agreement among participants is not essential at this stage. When all of the labels are recorded, go back over them, eliminating synonymous and combining highly similar ones. Then ask the participant responsible for each label to provide a definition for that label. At this stage, it is the responsibility of the person suggesting the definition to obtain agreement from the group. If such agreement cannot be obtained, the definition and/or label may be too ambiguous and should be dropped.

It might be best to hold meetings for supervisors and meetings for patrol officers separately at this stage since the supervisors may have a suppressing effect on the deliberation of the patrol officers. If separate meetings are possible, it would be extremely valuable to have each group react to the suggestions of the other group, perhaps at a joint meeting. This technique often initiates valuable discussion between supervisory and subordinate groups about the nature and definition of good performance.

When participants accept the categories and definitions of performance, those measurable by other means should be eliminated from consideration in developing the judgmental system. For example, rating schemes often include a scale labeled "attendance," a performance category best measured from personnel records, not judgments. The identified categories of performance should be carefully restricted to those best measured by judgments.

In the identification and definition phase, the behavioral aspects of each performance category should be emphasized; the participants must be encouraged to define each of the performance categories in terms of behaviors involved, rather than in global or nonspecific terms.

2. Criteria Development: Determine the behavior that will be used as an anchor to describe the various levels of performance for each performance category

This goal might be best achieved by combining a group meeting with individual efforts. Approximately ten supervisors and ten patrol officers might gather in separate meetings to discuss the performance categories and definitions. The participants can be told of the results of the first meeting, presented the definitions and categories, and encouraged to discuss these definitions until they have become fully familiar with their meanings. They can then be instructed to write examples of behaviors to

fit these definitions. (This is labeled "Item Generation" in Figure 6.) If some examples are provided, the participants will probably work more efficiently. They can then continue the procedure on their own for a week or two and return their examples to the personnel department at a particular time. The participants should be encouraged to work independently of one another and try to think of behaviors at all points on each of the performance categories (e.g., to think of three examples of behavior which are good, average, and poor for each category). If there are five categories and ten people each writing nine examples, the result should be 450 examples from each group. More realistically, there will be 150 to 200 usable examples from each of the two groups.

The usefulness of these items can be checked through a process of "re-translation," as described in the earlier section of this monograph.

3. Criteria Selection: Choose items which will act as anchors for the final rating scales

Once there is a pool of items for each performance category, items for actual use must be chosen. Because a desirable performance rating scale should clearly indicate to a supervisor the performance level implied at each point on that scale, a group of officers and supervisors should order the items in each category. The participants might simply assign a number from 1 to 10, along a scale from highly effective to highly ineffective behavior, to represent the level of performance described by the particular item. Items assigned very different positions have little value and should be discarded because they represent a disagreement concerning the specific definition of performance.

4. Choose a Format for the Performance Appraisal

There are at least three ways to put the scale together now that the level of performance for each behavioral description has been determined:

- a. Arrange the items in the form of a behaviorally anchored scale, with the descriptions arranged next to their assigned value on a vertical scale. In this case, the supervisor using the rating procedure might be requested to use one of a limited set of numbers on the scale itself to indicate the rating (e.g., pick a number from 1 to 9 which represents the performance of the individual). (See Figure 5)
- b. Arrange the items in five boxes representing five levels of performance from "inadequate" to "excellent," giving each box a number from 1 to 5 (much like the traditional graphic rating scale). The supervisor might then be asked to assign as a rating the number of the box which best describes the behavior of the ratee. (See Figure 4)
- c. Arrange the items in a check list format. The supervisor might then be asked to check those behavior descriptions that

adequately describe the person being rated. Because the scale values of the items are known, an average performance score can be calculated for each officer for each performance category. (See Figure 1)

Consulting the supervisors and patrol officers concerned about the final format will help insure its success. Although any of the three formats described above might be suitable, any clear preferences of the users should not be ignored since these personnel determine the final acceptance of the system.

Maintenance of a New Performance Appraisal System

Up to this point, the process of improving a performance appraisal system has been treated as if the problem concerns only moving from Condition A (use of an old system) to Condition B (use of a new system), implying that Condition B solves all problems. This is not the case. Two additional reasons necessitate continuing a program of development in the performance rating system. First, organizations and situations change, and a performance appraisal system must change with them. Plans should be made to examine the rating system regularly, perhaps once a year, to insure that it continues to function as intended. Officers, formed into groups to evaluate the adequacy of the system over the past year, might suggest the addition of new dimensions, or the expansion of some old definitions, or the addition or deletion of some examples of levels of performance.

A second, more subtle, yet equally important reason for continuing a program of development is the involvement of personnel who were not part of the original development procedure. The periodic evaluations of the system might be used for this purpose in such continued involvement. Unless involvement of all personnel is maintained, after two or three years perhaps, new members of the department might try to entirely change the system again. It is simply inefficient to change a personnel system completely every third year; it is more productive to insure regular evaluation and modification of a basically good and well-developed system through scheduled reviews planned, perhaps, by line officers.

A personnel system, whether a performance appraisal system or a selection system, should serve the needs of the department. It should not be so rigid and inflexible that the members of the department must change to fit the characteristics of the personnel system.

REFERENCES

- Dunnette, M.D. and Motowidlo, S.J. Development of a Personnel Selection and Assessment System for Police Officers for Patrol, Investigative, Supervisory and Command Positions. Minneapolis: Personnel Decisions Inc., 1975.
- Guion, R.M. Personnel Testing. New York: McGraw-Hill, 1965.
- Landy, F.J. "Performance Appraisal," in Police Personnel Administration, ed. O. Glenn Stahl and Richard Staufenberger (Washington: D.C.: Police Foundation, 1974).
- Landy, F.J. "The Validity of the Interview in Police Officer Selection," Journal of Applied Psychology, 61, 1976, 193-98.
- Landy, F.J., and J.F. Farr Police Performance Appraisal: Technical Report (University Park, Pa.: Law Enforcement Assistance Administration, 1975).
- Landy, F.J., and R.M. Guion "The Development of Behaviorally Anchored Scales for the Measurement of Work Motivation," Organizational Behavior and Human Performance, 5, 1970, 93-103,
- Terman, L. "A Trial of Mental and Pedagogical Tests in a Civil Service Examination for Policement and Firemen," Journal of Applied Psychology, 1, 1917, 17-29.

Appendix A

Sample Employee Performance Report

METROPOLITAN DADE COUNTY—PUBLIC SAFETY DEPARTMENT EMPLOYEE PERFORMANCE REPORT FOR POLICE OFFICERS ON ROAD PATROL

Name	Period Covered	Civil Service Status	If Prob., Date Ends	District
(last) (first) (initial) (to)				
Job Knowledge — Awareness of procedures, laws, and court rulings, and changes in them. (9) — (8) — (7) — (6) — (5) — (4) — (3) — (2) — (1) —	Judgment — Observation and assessment of the situation and taking appropriate action. (9) — (8) — (7) — (6) — (5) — (4) — (3) — (2) — (1) —	Initiative — Individual personal performance conducted without either direct supervision or commands, including suggestions for improved departmental procedures. (9) — (8) — (7) — (6) — (5) — (4) — (3) — (2) — (1) —	Dependability — Predictable job behavior, including attendance promptness, and reaction to boredom, stress and criticism. (9) — (8) — (7) — (6) — (5) — (4) — (3) — (2) — (1) —	

Demeanor — Professional bearing as determined by overall neatness of uniform, personal grooming, and general physical condition. (9) — (8) — (7) — (6) — (5) — (4) — (3) — (2) — (1) —	Attitude — General orientation toward the law enforcement profession and the department. (9) — (8) — (7) — (6) — (5) — (4) — (3) — (2) — (1) —	Relations with others — Ability to deal with people he comes into contact with during the performance of his job, including the public, fellow officers, and supervisory personnel. (9) — (8) — (7) — (6) — (5) — (4) — (3) — (2) — (1) —	Communication — Ability to make oneself understood and gather and transmit information, both in oral and written fashion. (9) — (8) — (7) — (6) — (5) — (4) — (3) — (2) — (1) —
--	--	---	---

Rater's Recommendation (for employees under consideration for a merit raise or permanent status)
 This is to certify that the overall performance of the subject employee _____ is _____ is not satisfactory.
 The employee _____ is _____ is not recommended for _____ a merit raise _____ permanent status.

This report is based on my observation and knowledge. It represents my best judgment of the employee's performance.

RATER _____ Date _____
 DIVISION _____ Unit _____

I have reviewed this report. It represents the facts to the best of my knowledge. I concur in the recommendation, if any, as to merit raise or permanent status.

REVIEWER _____ Date _____
 In signing this report I do not necessarily agree with the conclusions of the rater. I understand that I may write my comments on the reverse side. I have received a copy of this report.

EMPLOYEE'S SIGNATURE _____ Date _____

(Rater's summary comments should be put on attached sheets).

EMPLOYEE

Appendix B

Instructions and Examples for Performance Description Scales in Dade County

DADE COUNTY PUBLIC SAFETY

INSTRUCTIONS AND EXAMPLES FOR PERFORMANCE DESCRIPTION SCALES FOR ROAD PATROL OFFICERS

INSTRUCTIONS

The purpose of a rating scale is to provide an objective way of evaluating the "intangibles" of work performance. The rating scales which follow have been designed by your fellow supervisors to help you rate the performance of your officers on eight areas which have been consistently identified as important aspects of patrol officer performance.

The rating process is simple. Consider the individual to be rated and give a rating on each of the eight (8) aspects of performance. You will notice that each of the eight aspects is defined for you. You will rate each officer by describing how well the officer typically performs the job aspect. When rating the officer, keep in mind the definition of the aspect being rated.

In other rating systems, it has been found that simply defining the work aspect for the rater does not provide enough information for accurate and reliable rating. Additional information is needed about the points along the rating scale. You will find this additional information starting on page 4 of this booklet.

Look at the first set of scales entitled job knowledge, you will see that a series of examples of job performance related to Job Knowledge are presented. These examples are placed in three broad categories: high, average, and low. These examples have been provided by first-line supervisors of patrol officers throughout this department. The only examples provided are those which large numbers of supervisors agreed to put in a particular category (high, average, or low). We have found that these job performance examples help the supervisors make a more objective estimate of how well the officer has performed a particular job aspect during the rating period. We have also found that when supervisors use this kind of rating procedure, there is less confusion about the meaning of terms like "satisfactory," "outstanding," "needs attention," etc. which are usually found on rating scales.

You will notice that the rating scale itself is made up of a vertical line which has been numbered from 1 through 9. The purpose of the job performance examples starting on page 4 is to tell you exactly what level of performance is indicated by the various points along the scale. In other words, we are trying to give you an idea of the behavior represented by a 7 or a 4 or a 2 on the scale.

The performance examples which have been provided by first-line supervisors of patrol officers may not adequately describe the performance of the individual patrol officer you are rating. For example, some raters find themselves disagreeing with the HIGH descriptions because they expect that level of performance from all their officers. Consequently, they think that the descriptions should be considered AVERAGE. If you require that all of your officers consistently preserve all potential evidence at the scene of a crime, this does not make the behavior average; IT MAKES YOUR REQUIRED LEVEL OF PERFORMANCE HIGH. In law enforcement, there are many areas in which nothing short of excellence or HIGH behavior is acceptable.

Steps in Rating an Officer

The actual rating procedure should be as follows:

1. Turn to page 4 of this booklet labelled Job Knowledge: consider the definition of this aspect of performance; consider the examples of job performance which describe high, average, and low Job Knowledge, and decide on how much Job Knowledge the officer being rated has shown in the rating period.
2. Circle the point on the rating scale which best represents the amount of Job Knowledge demonstrated. This point can be any one of the hash marks along the scale (any one of the nine numbers or one of the halfway points between two numbers).
3. Turn to page 5, Judgment, Scale Two, in this booklet and rate the officer in a similar manner. Continue the process until the individual has been rated on each of the eight rating scales.
4. On unattached sheet headed "rater's comments" provide a written synopsis of the reasons for your ratings. You may choose to give your own examples of behavior that you actually observed from the ratee during the rating period.

(4)

Job Knowledge - Awareness of procedures, laws, and court rulings and changes in them.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "High" on "Job Knowledge" by supervisors.

High

Always follows correct procedures for evidence preservation at the scene of a crime.

Is fully aware of recent court rulings, and conducts himself accordingly.

Searches a citizen's vehicle with probable cause, thereby discovering smuggled narcotics.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Average" on "Job Knowledge" by supervisors.

Average

Arrests a suspect at 11:00 P.M. on a warrant only after insuring that the warrant had been cleared for night service.

Distinguishes between civil matters and police matters.

Seldom has to ask others about points of law.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Low" on "Job Knowledge" by supervisors.

Low

Is consistently unaware of general orders and/or departmental policy.

Arrests a suspect for a misdemeanor not committed in his presence.

Misinforms the public on legal matters through lack of knowledge.

Judgment - Observation and assessment of the situation and taking appropriate action.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "High" on "Judgment" by supervisors.

High

Calls for assistance and clears the area of bystanders before confronting a barricaded, heavily-armed suspect.

Notifies potentially dangerous situations before anything actually occurs.

Remains in his position and discontinues a high-speed chase before entering areas of high vehicle and pedestrian traffic, such as school areas.

Average

Issues warnings instead of tickets for traffic violations which occur at particularly confusing intersections for motorists.

Permits traffic violators to explain why they violated the law and then decides whether or not to issue a citation.

Does not leave a mother and daughter in the middle of a fight just because no law is being violated.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Average" on "Judgment" by supervisors.

Low

Enters a building with a broken door window instead of guarding the exits and calling for a backup unit.

Does nothing in response to a complaint about a woman cursing loudly in a restaurant.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Low" on "Judgment" by supervisors.

Continues to write a traffic violation when he hears a report of a nearby robbery in progress.

(6)

Initiative - Individual personal performance conducted without either direct supervision or commands, including suggestions for improved departmental procedures.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "High" on "Initiative" by supervisors.

High

Makes a special effort to find burglaries on his beat by carefully inspecting for signs of possible break-ins before the owners open their stores for business.

Comes to work early in order to check on the previous day's activities.

Seeks information about recent court rulings during his off-duty hours so that "good" arrests won't be lost by his actions.

The description to the right is an example of behavior of individual patrol officers who are usually rated "Average" on "Initiative" by supervisors.

Average

Fails to recognize and correct his own deficiencies without prompting from others.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Low" on "Initiative" by supervisors.

Low

Rarely checks the files for a suspect's friends and favorite hangouts.

Relies on his supervisor to make most of the important decisions for him.

Dependability - Predictable job behaviors, including job attendance, promptness, and reaction to boredom, stress, and criticism.

High

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "High" on "Dependability" by supervisors.

Remains cool under any circumstances.

Follows instructions.

Always gets to the station in time to check the daily log.

Average

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Average" on "Dependability" by supervisors.

Only uses a minimum of sick days each year.

Reports for duty even though he has a cold, if trouble is anticipated in the city.

Low

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Low" on "Dependability" by supervisors.

Panics upon receiving an emergency call.

Uses the condition of his squad car as an excuse to avoid responding to a call.

"Cracks up" in tense situations and shouts at other officers.

(8)

Demeanor - Professional bearing as determined by overall neatness of uniform, personal grooming, and general physical condition.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "High" on "Demeanor" by supervisors.

High

Is meticulous about personal hygiene.

Works to keep himself in shape even though he's 45 years old.

Wears a clean, pressed uniform.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Average" on "Demeanor" by supervisors.

Average

Polishes boots and brass every day.

Cleans out his squad car at the end of a shift.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Low" on "Demeanor" by supervisors.

Low

Wears a uniform with holes in it.

Gets so fat that he can no longer do his job properly.

Reports for duty with his hair uncombed and an obvious hangover.

Attitude - General orientation toward the law enforcement profession and the department.

High

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "High" on "Attitude" by supervisors.

Considers law enforcement a career, not just a job.

Takes part in a study of police officers' opinions being conducted by a local college.

Average

The description to the right is an example of behavior of individual patrol officers who are usually rated "Average" on "Attitude" by supervisors.

Seldom gripes about departmental procedures.

Low

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Low" on "Attitude" by supervisors.

Refuses training because he already is an expert.

"Goes out of his way" to defy departmental regulations.

(10)

Relations with others - Ability to deal with people he comes into contact with during the performance of his job, including the public, fellow officers, and supervisory personnel.

High

Takes the time to carefully answer a Rookie's questions.

Maintains friendly relations with the civilians in his patrol area.

Establishes good relations with the youth in his patrol area by answering their questions and letting them look at his car and equipment.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "High" on "Relations with others" by supervisors.

Average

Is considered "one of the boys" on his watch or shift.

The description to the right is an example of behavior of individual patrol officers who are usually rated "Average" on "Relations with others" by supervisors.

Low

Always has fellow officers riled up by his actions and remarks.

Uses racially-toned language in front of minority group members.

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Low" on "Relations with others" by supervisors.

Communication - Ability to make oneself understood and gather and transmit information, both in oral and written fashion.

High

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "High" on "Communication" by supervisors.

Turns in reports which are neat, accurate, and well written.

Carefully separates opinion from fact in his written and oral reports.

Uses correct grammar, spelling, and punctuation in written reports.

The description to the right is an example of behavior of individual patrol officers who are usually rated "Average" on "Communication" by supervisors.

Never has to be asked to repeat himself over the radio.

Low

The descriptions to the right are examples of behavior of individual patrol officers who are usually rated "Low" on "Communication" by supervisors.

Includes far too much trivial, irrelevant information in his written reports and radio communications.

Uses "choppy," incomplete language in his written reports.

Talks so fast over the radio that he is unintelligible.

