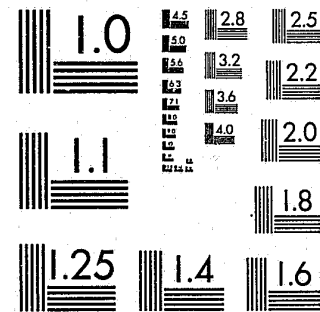


National Criminal Justice Reference Service

ncjrs

This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

National Institute of Justice
United States Department of Justice
Washington, D. C. 20531

4-23-82

80939

78-NI-AX-0056

Police Foundation

MF-1

X
IMPROVING POLICE PERFORMANCE MEASUREMENT:
IS THE TOPIC WORTH ONE MORE PAPER?

MARY ANN WYCOFF
PROJECT DIRECTOR
POLICE FOUNDATION
2564 BRANCH STREET, SUITE 1
MIDDLETON, WISCONSIN 53562

Prepared for presentation to the National Conference of the American Society for Public Administration, Detroit, April 12-15, 1981.

Portions of the perspective expressed in this paper developed during the course of a grant from the National Institute of Justice, Department of Justice (78-NI-AX-0056).

U.S. Department of Justice
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this copyrighted material has been granted by

Public Domain
National Inst. of Justice

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the copyright owner.

IMPROVING POLICE PERFORMANCE MEASUREMENT:

IS THE TOPIC WORTH ONE MORE PAPER?

Introduction

A few weeks ago, realizing it was past time to get serious about this presentation, I sent Steve Mastrofsky a note declining the opportunity to appear. In addition to having re-read the significant written discussions of police performance measurement, I had just read a draft of a book by Whitaker, Mastrofsky, Ostrom, Parks and Percy entitled Measuring Police Agency Performance, and was convinced that everything that needs to be said on the subject has been said at least once. Does the issue really require one more paper?

I thought not if the focus was, indeed, to be on improving the quality of available data. A colleague, Bob Petersen, once asked in a staff meeting: "is that which is not worth doing, worth doing well?" If the question is applied to most existing police performance data, my answer is "no".

Finally, one can ask whether there is any reason to write one more paper on the subject when there seems to be little response to those already produced.

Despite those misgivings, I am here -- primarily because Steve does not easily accept "no" as an answer and because this forum seems like a suitable one in which to argue for new directions in work on police performance measures.

Having stated so baldly the reservations about our original topic, it might be useful to attempt to justify them briefly.

1. Everything worth saying has been said. In 1978 Gary Marx published "Alternative Measures of Police Performance" in which he reviewed limitations

of traditional measures and proposed several exciting alternatives. This stands as one of the most significant publications on the issue of performance measurement. During the last fifteen years a critical body of literature developed which demonstrated the problems of official crime rates (Biderman and Reiss, 1967; Black, 1970; Cook, 1977; Maltz, 1975; Seidman and Couzens, 1974; Skogan, 1974; Wolfgang, 1963), clearance rates (Greenwood, et.al., 1977; Hatry, 1975; Skolnick, 1966), arrest rates (Greenwood, et.al., 1977; LaFave, 1965; Sherman and Glick, 1980) and other agency-produced data (Kelling, 1976). In 1978, George Kelling argued for "...a modest moratorium on the application of crime related productivity measures (p. 205) until it could be determined what police actually do and how those activities can be measured. In 1979 Wycoff and Susmilch argued that knowledge about police impact would depend on the development of direct (as opposed to "surrogate") measures of police performance. Similarly, Kelling, Wycoff and Pate (1980) argued that measurement of "inputs" and "outcomes" would produce only an un-illuminated "black box" perception of police impact and that substantial attention would have to be given to measures of process, measures of what police do (i.e., "performance"). Similar points were made again by Wycoff, Susmilch and Eisenbart (1980) who additionally discussed limitations in the use of general citizen surveys as indicators of performance. Finally, in their forthcoming work, Whitaker, Mastrofsky, Ostrom, Parks and Percy summarize some of these earlier points and make very powerfully the case against general measures of police performance. Measures must be specific to function and the choice of measure will have to change as policing needs of a community change. They further point out that the selection of performance measures has seldom been guided by theory and that a great deal remains to be done in the area of theory development. They argue effectively that performance measures

be used as learning tools. With even this brief nod to the literature, I am convinced that there is nothing important left to say except, let's get on with the hard work of hammering out measures and methods for gauging actual performance.

2. That which is not worth doing is not worth doing well. If the interest is in improving the collection, storage, and retrieval of existing agency data for the purpose of testing hypotheses across agencies, I cannot share the interest. I think such hypothesis testing at this time is not methodologically feasible and is potentially misleading. It is not methodologically possible because, no matter how great the improvement within individual agencies in terms of data collection and storage, there will always be agency and community differences which will make virtually impossible to know whether indicators which are named the same by various agencies actually mean the same across the agencies. The way in which crimes are reported and recorded, the way in which arrests are designated and clearances determined, for example, can all vary across agencies. Given these differences, one cannot know whether quantitative differences in indicators across agencies represent differences in actual behaviors (e.g., incidents of crime) or differences in ways in which data about the behaviors are acquired.

As just one example: in one agency in which the Police Foundation worked, it was determined that reports of robberies or burglaries from apparently intoxicated callers were not recorded officially until detectives had verified that there was good reason to believe a crime actually had occurred.¹ Had we compared the number of such incidents in this agency

¹We did not determine the extent to which the calls were followed-up by detectives.

which was conducting a program to reduce robberies to the number of such incidents in a city without a program to reduce robberies and without a program for screening intoxicated complainants, we might have concluded falsely that these data indicated the effectiveness of the special robbery program. Had we compared the department with the screening process to one without in terms of the number of arrests relative to all recorded robbery incidents or in terms of the number of clearances relative to recorded incidents, we might have concluded that the department which screened complaints was more effective in "solving" robberies. In reality, it might have been more effective in reducing the number of reported robberies which it put on the books.

This is an example of only one agency difference; there most probably is an almost "unknowable" number of such differences which will serve to make apparent differences among agencies uninterpretable. The number is "unknowable" because the conditions affecting the differences are likely to be changing over time and because almost no researcher who gathers data from several agencies will have the resources even to begin to examine the various processes for coding behaviors and recording data in these agencies.

In short, it probably is not worthwhile to improve agency records for the purpose of doing comparative evaluations of agencies "effectiveness" or "outcomes" or to test hypotheses across agencies about the effects of different types or levels of inputs.

Cross-agency comparisons may be useful, however, if the comparison is for the purpose of learning about agency processes (Whitaker, et.al., 1980) rather than for the purpose of testing hypotheses or evaluating agencies. Differences among agencies may suggest questions to be asked

about differences in community characteristics, data collection methods or methods of operation. The statistics serve then as guides to additional directions in research; they may be very useful for suggesting questions while nearly useless for providing answers.

Some agency data can also be useful for assessing an agency's current practice relative to past practice and for this type of purpose, an agency may find it very worthwhile to seek ways of improving data content, reliability, storage, retrieval, etc. However, the types of improvements to be made may be largely peculiar to the agency, depending on its measurement objectives and on the types of problems currently existing in the data collection and retrieval methods. There are individuals and organizations (e.g., P.E.I. and I.A.C.P.) who can help police agencies identify their data problems and help formulate solutions. Although this may be an issue of great importance to many administrators who are present, the breadth of possible problems and solutions is greater than the limits of this paper and outside the realm of my own interest.

It is heretical for a researcher to proclaim disinterest in improving data sources. In defense I can only say that my disinterest reflects a disinterest in the types of data police agencies tend to store which, for the most part, are data which reflect primarily the "outputs" rather than the "processes" of police response to problems. Looking at outputs leaves us, again, with the rather uninteresting black box in our hands. It is the process, not the outputs and not the reactions to the process or its outputs or reactions to them, which should be captured by performance measures.

Random House Dictionary defines "perform" as "to carry out; execute, do." A performance can be defined simply as the "execution of work". In much of the literature on performance measurement, performance seems presumed to

connote an evaluative component. Perhaps because the interest in performance measurement has grown out of an interest in evaluation, there has been an almost automatic leap to the conclusion that measurement of performance means the measurement of "outputs", "outcomes" or "effects". It may be this that has led us into the "black box" in which inputs are linked in uninterpretable and untheoretic ways to outcomes. The theories can't be constructed because there is so little knowledge about the linkage variables in the model. The performance is the behavior or the process which links the input to the output. It is the individual or agency effort or work, independent of the consequences of it or the reaction to it. A good performance may be a desirable organizational product or outcome, but the nature of the performance as objectively described, is independent of the reaction to it or the effect it has on other conditions or behaviors.

Most of the types of measures which have been labeled "performance measures" have been measures of presumed effects of behaviors or of reactions which are assumed to be produced by the behaviors. Very rarely has the performance itself been measured. This has been true whether the interest has been in measuring police effectiveness in dealing with crime, traffic, domestic disputes, drunks, other disabled persons, frightened persons, inquisitive persons, crowds, mentally disturbed persons, juveniles, lost children, hazardous conditions, or whatever. We attempt to assess the effect of some input variable (e.g., dollars, personnel, technology, organizational structure) on an outcome without having any idea what the police actually did in the performance or process which linked the input to outcome. If a correlation indicates that something "worked", we don't know what or why; if the lack of a correlation indicates that something didn't work, we still don't

know why. For all of the measurement and analysis, little of value is learned. We learn nothing which will lead to the development and testing of theories and we learn nothing about the operation and implementation of programs which can be passed on from the agency studies to an agency which might wish either to adopt or adapt a program or policy.

In short, existing data are uninteresting because they do not measure performance. But there is another reason why I lack motivation to improve existing data: for the most part, they reflect only the outputs of the crime-focused police functions. Focusing on crime functions -- no matter how well relevant performance might be measured -- leads to only a limited understanding of the totality of police performance. Numerous studies demonstrate that, either in terms of what police are asked to do by citizens or what they actually spend their time doing, crime-related issues account for a small part of total police patrol² effort. Concentration on improvement of crime-effectiveness data draws attention and resources away from the development of performance measures relative to all the other functions which a police agency performs for the community. Counting that which can be easily counted undoubtedly has contributed to the "police-as-crime-fighters-myth" (Manning, 1977). It comes to be believed that whatever is measured and discussed must be important; if crime efforts are the only ones which are counted and reported, they must be the only ones which matter. If an agency counts the number of arrests a patrol officer makes but does not count the number of that officer's community contacts or the number of disputes

²There are no measures of total agency effort in various types of functions.

resolved without arrest, then that officer may reasonably come to assume that arrests are the most important product to be delivered. Can it be very surprising then, if researchers compare police agencies only in terms of crime rates, arrest and clearance rates, that police agencies and the public come to assume that these are the only meaningful products of police agencies? Do we then have any right to lament that the public has no understanding of the range of the police role? Public blindness about the nature of the police job in this country can't be dropped on only the doorstep of the television producers; researchers have contributed to the miasma.

Crime-related data have been used by researchers not always because they were believed to represent the most or the only important indicator of police performance, but often because they were available, affordable and could be easily acquired from a number of agencies. By using and re-using the figures, debating and re-debating the findings based on them, researchers have contributed to the belief that crime-related performance constitutes the only meaningful basis for either agency self-evaluation or cross-agency comparisons.

In the process of doing crime-related analyses, we got better at what we did. We became much more sophisticated in our analyses and we did become more insightful about the problems of the data. Those are positive steps, both for researchers and police agencies, but while we were producing more and more paper about how to measure crime-related performance in better ways, we contributed to the belief (without examining it) that this was the most important police function to be assessed.

I regret that this has happened and personally wish to contribute no further to this development. Since that is a rather dramatic statement,

let me be quite clear about what it does and does not mean.

a. It means that I do not want to turn my own attentions to improving the quality of crime-focused data, except insofar as individual research projects require it.³

b. It does not mean that measures of crime-effectiveness are to be considered unimportant. Obviously, dealing with crime is a critical police function which should be examined critically. Work conducted toward this end is valuable and the argument is not that all such efforts toward this end should be halted. However, such work should be counter-balanced by efforts to measure police effectiveness on other dimensions and some re-researchers should be focusing their efforts accordingly. The argument being made is for righting an imbalance which probably has contributed to a narrow and unhealthy view of the police role.

Finally, I questioned whether another paper on police performance was worthwhile since:

3. No one seems to listen, anyway. Despite the fact that the cautions against doing so are not new, respected scholars continue to produce articles based on surrogate measures and to aggregate crime-related data across jurisdictions without examining or questioning its comparability. This happens because time and financial resources do not support the development, testing, and implementation of alternative measures of performance.

³In an evaluation we have completed and in one we are now developing, quantitative measures of arrests are augmented with measures of prior arrest records of arrestees and with data about case outcomes.

Among practitioners, a progressive chief can't give attention to alternative performance measures and reward structures because the nature of the agency requires that it first learn to count those things which can be easily counted. Crimes and crime-related responses will be counted first.

From Washington comes a new pledge to do something about crime-in-the-streets. This is a loud wind beginning to howl which will make it difficult to hear voices discussing the need to measure non-crime-focused police behaviors.

But perhaps this is overly pessimistic; perhaps it is simply too early to expect results. Chief R. F. Lunney (1978), a Canadian police administrator, has noted that

The implementation of a performance indicator system within a department must be regarded as a long-term program. As with all opportunities in the system age, results must be preceded by an investment of resources, almost always considerable, in the process of "tool building". (p. 9).

So, someone may be listening; it may simply take time for them to respond to the message and to implement changes. The fact that it may become increasingly hard for the message to be heard may be the primary rationale for repeating it. Although foreign travelers don't always have luck with this method, bureaucrats are sometimes successful in getting the public to accept strange-sounding words if only they are repeated with sufficient frequency. So, one more time the message: Let's not spend all our energies on the attempted improvement of existing data about the crime function. There are, according to the American Bar Association (1971) at least six other functions. Let's begin to determine ways of measuring the performance and outcomes of these.

Which comes first? Since the arguments in this paper tend to be based, in part, on gloomy predictions of every limited resources, where do we put our first efforts -- on the development of process/performance/behavior measures or on the development of outputs of non-crime police functions? There are two responses one can make:

a. The lack of performance or behavior measures is a severe handicap when trying to assess any type of police function. From the standpoint of theory building, the identification of the process or performance variables should precede the identification of output and outcome variables since only through an understanding of the process would it be possible to anticipate the variety of possible products, effects, or reactions.

b. For many non-crime functions, and frequently in the case of crime-related functions as well, the quality of the performance may be the most significant output. In developing measures of process, we will simultaneously be developing means of assessing certain types of outputs.

If a choice is to be made, it seems preferable to begin to struggle with measuring process as performance, with measuring what police do.

An example of performance: In order to be more concrete about what is meant by measuring performance, we can consider what could be known about an officer handling a call, in this case a citizen's report of a household burglary.⁴

⁴The example is of a patrol officer, but performance measures could be developed for any unit or type of assignment. Additionally, it is the work of the patrol officer which probably has the greatest direct effect on the community. And, yes, this is a crime-focused example, chosen for lack of imagination in thinking of an easier one to consider!

"Performance", of course, masks a host of behavioral dimensions and the choice of dimensions to measure and the elaborateness with which they are measured may reflect both one's interests and one's resources. A police administrator or a researcher interested in documenting patrol officer response to any citizen's call might minimally want to know:

1. what activities the officer conducted (for this there could be essentially a check list of the steps which could be taken in handling particular types of calls);
2. how long it took the officer to arrive at the scene and how long he or she stayed;
3. how the officer treated persons at the scene.

From this information it should be possible to determine both the content and the quality of the officer's performance. Content simply refers to the type of activities the officer performs. For example, in the case of the household burglary, the officer might:

- determine the identification of anyone residing at the residence,
- examine the crime scene and secure it for any additional investigative work,
- ask questions about property stolen (identify), circumstances surrounding the incident, persons living or working in the neighborhood who might have been witnesses,
- write a report,
- discuss a report,
- discuss with victim procedures for filing insurance claims and for increasing physical security to deter possible future burglaries,
- discuss the probable developments in the case,

- mention to the victim possible psychological reactions to household burglaries,
- give name and phone number of appropriate police person to victim.

The combination of these activities would constitute the content of the performance. The quality of the performance would be determined by an assessment of the officer's competence and style. The level of competence would reflect the extent to which an officer did what he/she was trained to do. Style would reflect the interpersonal manner the officer used while conducting the activities. A performance might be conducted competently by an officer who did everything required by the department and yet be conducted in either a positive or negative style, depending on whether the officer is civil and polite or uncivil and rude.

Separating the concept of performance into those of competence and style would allow for a test of the hypothesis that "it's not what you do but how you do it that counts" and would allow a determination of which types of styles are more or less acceptable to which publics.

Individual or agency performance? The definition and example of performance used here clearly relate to the performance of the individual patrol officer, leaving open the question of indicators of agency performance. While it will be argued that measures of individual performance can/should be combined to produce measures of unit and agency performance, it is certainly true that such an aggregation cannot capture all important aspects of an agency's performance. Aggregating the performance of individual communications personnel will not indicate how many calls to the department go unanswered or remain for an unacceptable amount of time in an answering que. (Examination of individual performances of communications personnel might, however, help

explain why these conditions existed.) The argument here is that there are few, if any, indicators of agency performance which would not need to be explicated by examination of individual performance and that, therefore, both types of indicators are necessary. Further it is argued that much of what does matter about agency performance can be measured by aggregations of individual performances; indeed, can be measured only by such aggregation. The types of services delivered to the community and the way in which they are delivered constitute the essence of policing; measures of performance which do not assess content and quality of performance simply miss the point, and while they might provide researchers with a quick (if hollow) comparative fix on organizations, they will provide organizations no information about what, how or whom to change in order to improve service delivery. Aggregation of appropriate individual measures can indicate whether different parts of the community are receiving different types and styles of treatment, whether particular types of calls or geographical areas of the city are more problematic for officers, whether particular units consistently perform better or worse than others, whether unit performance is a function of a minority or a majority of unit members, whether particular squad or division commanders appear more effective than others, etc.

A critical question that aggregation of individual performance measures could not answer is that of the distribution of resources and services according to need and demand. (See Whitaker, et.al., (1980) and Wycoff and Manning, 1979.) Individual performance could provide only a hint as to whether there were too few officers available during a period of high demand. Measures of resources relative to measures of need (e.g., numbers of actual crimes, accidents, disputes, etc.) or demand (requests for service) are

indicators necessary for assessing the management of the organization. Strictly speaking, however, they are only quantitative measures of the availability of resources; taken alone they cannot indicate the amount, type or quality of the service actually delivered. The agency-level indicator, unsupported by measures of individual (or aggregated to unit) performance still leave only the black-box understanding of organizational operation.

Having acknowledged that both agency and individual level indicators are necessary, I retreat to the kind of argument made previously: I will leave the development of the one type of indicator to others and put my energies into the one which seems to me more critical. I consider the development of individual performance measures to be the most worthwhile because they can be used:

- as the basis for determining policy and training needs,
- as the basis for locating supervisory or command strengths and weaknesses in the organization,
- as the basis for rewards and punishments which can shape the role orientation of the organization (which is to say that good measures of an individual's actual performance would make it possible to reward non-crime related activities as well as crime-related activities).

In short, the individual measures provide the most useful information about ways of changing, shaping, and managing the organization.

Developing the measures. The only reasonable way to develop performance measures is through a joint effort of researchers and police practitioners. Researchers working alone will not be able to identify the organizationally

relevant indicators and will not necessarily know which types of data feasibly (politically and practically) can be collected within police organizations. Working within the constraints of a funded project for which performance indicators are needed, researchers will seldom have any choice but to measure that which can be measured most quickly and cheaply. Similarly, individual police practitioners or agencies will find it difficult to free the time and other resources necessary for the development of a new evaluation system. The effort must be a joint one, funded for a sufficiently long period to permit debate, design and testing of indicators, and supported for no purpose other than the production of performance indicators and methodologies. Perhaps such an effort is the ultimate objective of the Accreditation Project which Mr. McLaren will discuss. If not, I believe the work just completed by Whitaker, et.al. and the work of the Accreditation Project provide a proper spring board for a project on measures and methods.

While the undertaking might take a lengthy period of time, the development of performance measures and methods need not be anticipated as an un-ending task. It is the case that different communities may have different service needs and expectations and that the same community may have different needs and expectations over time. This need not mean, however, that every organization must develop and re-develop its own indicators as conditions dictate. It should be possible to develop a fairly standard set of indicators which could remain in place. What would change would be the organization's emphasis on particular contents and styles of performances; the indicators would not change, but the desired "reading" from the indicators might. While different types of performances might be desirable for different communities,

standard indicators would allow researchers to determine whether different types of performances were actually delivered in different cities and what relationship these differences might bear to indicators of other types of outcomes (e.g., citizen satisfaction, citizen fear, use of the streets, levels of undesirable activity, etc.). Indicators which provided for relatively objective reporting of behavior would provide the information for evaluating that behavior but would permit separation of the documentation of the behavior (performance) from the evaluation of it.⁵

An example of a methodology. Given the best of all possible worlds I would, as a researcher, always use trained observers to record performance but obviously that is an extremely time-consuming and expensive methodology. While it is a method which should be used in the testing of other methods, it is unlikely (at least in the current budget era) that many research projects will be able to support extensive observation. Nor is it likely that observation will be feasible (or even desirable) as a way for the police organization to conduct routine performance evaluations.⁶ In lieu of direct observation, I would like to re-advocate a method outlined by Gary Marx (1978) and reported by him as having been used by Baltimore and some other police agencies (p. 25).⁷

⁵ Again, there is a distinction between the evaluation of the performance as behavior and the assessment of reactions to it or presumed consequences of it. Measuring outcomes remains a task separate from that of measuring and judging performance.

⁶ The "best" measure of performance will be the one that serves researchers and police administrators equally well.

⁷ The method has also been used by Bordua and Tifft (1971) and by Wycoff, Brown and Petersen (1980).

A sample of citizens who have requested police service would be called and asked a series of standard questions about that service.⁸ The particular set of questions used could vary, depending on the nature of the call. The caller could determine how long the citizen perceived the response time to be, how long the officer was perceived as being at the scene, what actions the officer(s)⁹ took (perhaps from a check list similar to that suggested for documenting a burglary response, p.12), the manner of the officer's behavior toward the citizen, and the citizen's reaction (an "outcome" measure) to the officer's handling of the call.

Calls could be made on a completely random basis or certain types of calls, geographic areas, population groups, or officers could be purposefully sampled whenever problems or suggestions of problems indicated such a need. These telephone follow-ups could constitute a standard component of officer or unit evaluation and could be routinely conducted by groups of patrol supervisors who might periodically be assigned this duty.¹⁰ As Marx points out, no single interview should constitute an evaluation; enough interviews should be done for each officer so that a pattern of behaviors and responses could be determined.

⁸ Many agencies do not record requests for service which do not result in dispatching an officer to the scene. Evaluation of the agency (or of the performance of the telephone operators) should require that this information be stored and that these types of calls be followed-up, as well.

⁹ If more than one officer responds, the behavior may have to be documented as team behavior unless individual performances stand out to the citizen.

¹⁰ Or they could be conducted by a special inspections unit, by police telephone operators trained to conduct the interviews as relief from regular switchboard responsibilities, or by a completely external polling agency.

Such a system for performance measurement would seem to have advantages for police officers, police supervisors and administrators, the general public and certainly for researchers. Police officers could derive satisfaction from knowing that the full range of their responsibilities was documented and could serve as the possible basis for reward and recognition. Police supervisors and administrators would have a means of assessing the service actually being delivered in the field. The questions used could help the police communicate to the public a set of "reasonable" (from a police perspective) expectations for the police performance. Citizens would be given the opportunity to have input into the police organization without having to make a special and unusual (and perhaps uncomfortable) effort to do so. Researchers could hope to acquire good performance data quickly and inexpensively and could, at the same time, acquire outcome data in the form of citizen attitudinal responses to service.¹¹

Obviously such a system is not without its drawbacks. It would be more expensive than traditional "performance" evaluation systems. Such monitoring might very well be opposed by employee organizations unless the department could demonstrate that assessment would be based only on patterns

¹¹ Such an information system would be responsive to Bordua and Tift's (1971) argument that

The most basic change necessary for police departments is to conduct research or to create a departmental role which can discover and present detailed information concerning the effect of various police practices on community attitudes so that policy decisions can be made rationally and knowledgeably. The information can make possible decisions which weigh the "costs" and the "gains" of each practice or policy as it is actually operationalized on the street. (p. 157)

and that rewards could be provided for a greater variety of good performances than was previously true. Researchers would point out that citizens' perceptions of police performance will not always be objective. All of these things are true; the response is that the potential payoffs justify the cost and effort and that the data would be far superior to any performance data which is now routinely available. The subjective impressions of citizens are those which matter in forming community attitudes toward the police agency. Analyses of patterns of subjectivity, in combination with follow-up research, could help an agency determine whether styles of officer behavior should be changed, whether citizen expectations needed to be modified or whether both changes should be made.

The potential richness of the data and its utility for both inter- and intra-agency research and as an administrative tool would seem to justify serious examination of this methodology.

Summary

This paper has questioned the worth of giving any more immediate attention to the improvement of existing measures of police performance, since these measures tend to be measures of outputs rather than performance and to focus almost exclusively on the crime-related components of the police role. Using only the crime-related indicators has served to exaggerate the importance of that function relative to other important ones which police perform.

Attention now should be given to developing outcome indicators for non-crime functions and to developing measures of actual processes, behaviors, performances as distinct from reactions to, and presumed consequences of,

these performances. The greatest need is to fill the "black box" between "inputs" and "outcomes" that are assumed to be related by means of behavior linkages which we currently have no way of examining.

An argument has been made for a consortium of police researchers and practitioners to create indicators of behavior which could be used similarly across agencies, though perhaps gauged differently by various agencies as local needs/desires required. The reports of behavior (independent of the subsequent ratings or evaluations of the behaviors) could be used comparatively across agencies. They could be used internally to assess individuals and units and to determine policy, training and supervisory needs.

A methodology based on calls to citizens who have requested service is recommended.

REFERENCES

- BIDERMAN, A. D. and A. J. Reiss, Jr. (1967) "On Exploring the 'Dark Figure' of Crime," The Annals of the American Academy of Political and Social Science, 374, pp. 1-15.
- BLACK, D. J. (1970) "Production of Crime Rates," American Sociological Review, 35, pp. 733-48.
- BORDUA, D. J. and L. L. Tifft, (1971) "Citizen Interviews, Organizational Feedback, and Police-Community Relations Decisions," Law and Society Review, November, pp. 155-182.
- COOK, P. J. (1977) "Punishment and Crime: A Critique of Current Findings Concerning the Preventive Effect of Punishment," Law and Contemporary Problems, 41, pp. 164-204.
- GREENWOOD, P. W., J. M. Chaiken, J. Petersilia (1977), The Criminal Investigation Process. Lexington, Mass.:D.C. Heath.
- HATRY, H. P. (1975) "Wrestling with Police Crime Control Productivity Measurement," in J. L. Wolfe and J. F. Heaphy (eds.), Readings on Productivity in Policing. Washington, D.C.:Police Foundation. pp. 86-128.
- KELLING, G. L. (1976) "Developing Indicators of Program Effectiveness: A Process," a paper prepared for the Conference on Emergency Medical Services: Research Methodology, Atlanta, Georgia, September, 1976.
- KELLING, G. L. (1978) "The Role of Research in Maximizing Productivity," in P. Engstad and M. Lioy (eds.), Report of the Proceedings, Workshop on Police Productivity and Performance. Ottawa, Canada:Ministry of the Solicitor General of Canada.
- KELLING, G.L., M. A. Wycoff and T. Pate (1980) "Policing: A Research Agenda for Rational Policy Making," in R. V. G. Clarke and J. M. Hough (eds.), The Effectiveness of Policing. Westmead, England: Gower Publishing Company Limited.
- LaFAVE, W. L. (1965) Arrest: The Decision to Take A Suspect Into Custody. Boston:Little, Brown.
- LUNNEY, R. F. (1978) "The Problems of Definition and Measurement of Police Productivity," in P. Engstad and M. Lioy (eds.), Report of the Proceedings, Workshop on Police Productivity and Performance. Ottawa, Canada:Ministry of the Solicitor General of Canada.
- MALTZ, M. D. (1975) "Crime Statistics: A Mathematical Perspective," Journal of Criminal Justice, 3, pp. 177-194.
- MANNING, P. (1977) Police Work: The Social Organization of Policing. Cambridge, Mass.:Massachusetts Institute of Technology Press.
- MARX, Gary (1978) "Alternative Measures of Police Performance, in R. C. Larson (ed.) Police Accountability. Lexington, Mass:Lexington Books.
- SEIDMAN, D., and M. Couzens (1974) "Getting the Crime Rate Down: Political Pressure and Crime Reporting," Law and Society Review, 8, pp. 457-93.
- SHERMAN, L. W. and B. Glick (1980) "The Validity of Arrest Rates for Cross-Sectional Analyses, a proposal submitted to the Office of Research and Evaluation Methods, National Institute of Justice." Washington, D.C.:Police Foundation.
- SKOGAN, Wesley (1974) "The Validity of Official Crime Statistics: An Empirical Investigation," Social Science Quarterly, 55, pp. 25-38.
- SKOLNICK, Jerome (1966) Justice Without Trial, New York:John Wiley.
- WHITAKER, G. P., S. Mastrofski, E. Ostrom, R. Parks and S. Percy (1980) Measuring Police Agency Performance (draft of a report to be published). Chapel Hill, N.C.:Institute for Research in Social Science at The University of North Carolina.
- WOLFGANG, M. E. (1963) "Uniform Crime Reports: A Critical Appraisal," University of Pennsylvania Law Review, 11, pp. 708-738.
- WYCOFF, M. A., C. E. Susmilch (1979) "The Relevance of College Education for Policing: Continuing the Dialogue," in D.E. Petersen (ed.), Police Work: Strategies and Outcomes in Law Enforcement. Beverly Hills, Calif.:Sage Publications, Inc.
- WYCOFF, M. A. and P. K. Manning (1979) "Community Service Demand and Public Agency Response, a proposal submitted to the National Science Foundation." Washington, D.C.:Police Foundation.
- WYCOFF, M. A., C. E. Brown, and R. E. Petersen (1980) Birmingham Anti-Robbery Unit Evaluation Report (draft of a report to be published). Washington D.C.:Police Foundation.
- WYCOFF, M. A., C. E. Susmilch, and P., Eisenbart (1980) Reconceptualizing the Police Role (draft of a report to be published). Washington, D.C.:Police Foundation.

END