EVALUATION IN THE CRIMINAL JUSTICE SERVICES

W.Clifford

Australian Institute of Criminology

EVALUATION IN THE CRIMINAL JUSTICE SERVICES

by

W. Clifford

Published and printed by the Australian Institute of Criminology, 10-18 Colbee Court, Phillip, A.C.T., Australia. 2606

Clifford, William, 1918-.
Evaluation in the criminal justice services.

Includes bibliographical references.
ISBN 0 642 89179 6.

1. Criminal justice, Administration of - Evaluation. I. Australian Institute of Criminology. II. Title.

364

INTRODUCTION

Most people come to evaluation with an ulterior motive - whether they recognise it or not. They are looking for a reason, for an excuse or perhaps a justification. For some, evaluation is an instrument of power. Without having to reach the highest administrative levels, they find themselves allowed to pronounce on the projects and programmes of so many others : and there is something enticingly God-like about making judgements without actually carrying responsibility. That is why evaluators should not themselves be immune from evaluation. There are those seeking to establish the efficiency of the services in which they work or for which they carry responsibility. They may try genuinely to be detached and objective but we cannot be surprised if they lean towards the evaluative methods which justify their policies and reinforce their own preconceptions. Then there are the technically minded, the methodologically committed, the honest devotees of cost-benefit analysis. They really believe that the medium is the message, that the process is what matters and that values that cannot be quantified just get in the way.

These are caricatures of the varieties of interest in the modern art - or perhaps science - of evaluation. Few people fit neatly or exclusively into one or the other type: but collectively these types represent the unimaginative way in which

evaluation is frequently learned and frequently taught. It is a subject which can entice the unwary into dogmatics. It is, therefore, prudent to approach evaluation with a healthy dash of cynicism - with a clear warning of the constant need for a lively imagination i.e. for the important facility not always exercised to stand apart from the project or process of evaluation and to visualise the same "obvious" facts in a very different setting. Even the self-evident can be misleading if the context is too specific, and comparisons are unduly restricted. To take but one very striking example, most planners, administrators and/or economists are sold on the importance of conducting feasibility studies before investing in large scale construction projects. We need to know the inputs and the returns over a given period of years. We should know at what point we break-even and begin to show a profit - and we should be able to calculate this within a reasonable period of time. If this is not possible or if reasonable returns cannot be foreseen - and perhaps calculated - we should consider alternative uses of the same resources. Economic man sees all this as obvious good sense. But, had civilisation applied such good sense to its development over the years, a great deal that we now treasure could never have seen the light of day. Imagine, a feasibility study being conducted on most of our great cultural heritages, at the time they were being planned. The Pyramids of Egypt and the Taj Mahal for instance. Who could have justified

such enormous investments in the interests of the dead. Europe's great cathedrals or the ornate temples of Asia which are now the focal points of world tourism. Who, at the time of their conception, could possibly have looked into a future with millions of people flying around the globe. They must have appeared to any evaluators, years after they were built, as so many human follies, white elephants, as projects not justified by their own objectives. And to bring this nearer home and much nearer our own times, it is very doubtful whether the Sydney Opera House would have survived either impartial feasibility studies in the past or a thorough going economic evaluation today.

The element in any evaluation which changes facile conclusions is the element of time. Given enough time almost any investment in a project or programme may live long enough to bury its detractors. It may eventually show economic returns never suspected. We abandoned railways and tramcars with all the valuable fixed capital which had been provided for us when labour was cheap and when steel was more readily available. We abandoned these because of the greater profitability of road transport with cheap oil. Then the price of oil changed and we suddenly became aware that it was a scarce resource. Tramway systems emerged as the world's cheapest form of public

transportation : but the opportunity had passed : we had torn up the lines and new investments could not justify returns. The countries that kept the trains and the trams because they did not have the means to build new highways for the millions of new cars, which they could not produce anyway, now emerge as having been economically prudent. Consider how the evaluations would have read at different periods of time.

The fact that we cannot be omniscient, affects any kind of evaluation. In the investments made in the form of technical assistance to developing nations, the political situation has had to be taken as a constant by most of the teams engaged on feasibility studies or evaluations for the international agencies or for the developed nations providing most of the resources. Usually the law and order system, and particularly the propensities for corruption, had to be ignored in such exercises. The result has been so much wasted paper on models and forecasts and, more seriously, so much wasted in terms of resources for development. Yet, again, if all the money creamed off such investments by the revolutionaries or by the corrupt, eventually finds its way into other profitable enterprises - and if a long enough time span is brought into perspective then the "waste" of today could become the "benefits" of tomorrow.

The fact is that the word "evaluation" is economic in origin : and therefore it has restrictive connotations. The Concise Oxford Dictionary defines it as "finding a numerical expression for" and it is precisely in this meaning that its limitations lie. We have seen that numerical indicators are likely to vary with the amount of time and the amount of knowledge brought to bear. That is why we need imagination to climb outside of the mathematical or economic parameters of evaluation from time to time to get a more realistic appreciation. This is also necessary for evaluators to develop the humility they need for their work to become more truly meaningful. We have to remember not to value only what is measurable. An evaluators true vocation is to find measures for what is valuable, but not the reverse, this process in the preoccupation with techniques.

Having provided then all the caveats and cautions : having warned of the need to examine very carefully not only the data but the assumptions undulying the data collection and its interpertation, it is now possible to proceed to a consideration of the true worth of evaluation. For whatever its shortcomings, evaluation is a necessary and indeed useful way of assessing our policy making. The drawbacks we have already reviewed are not

arguments against evaluation - only arguments for making it more sensitive to the human condition which can never be appreciated in economic terms alone.

This is of special importance in relation to the criminal justice services. For, before beginning evaluation, we have to examine our own objectives. What are we trying to do? To eliminate all crime from society? That is not only impossible but a deliberate ignoring of the nature of crime. Perhaps, I need not pursue this beyond reminding you, that Christ was a criminal, that we now have too many laws for any citizen to void illegality always, and that our criminal justice systems are highly selective. It is no coincidence that more of the poor than the rich crowd our criminal courts. Ghandi was not far wrong when he declared that

> "Honesty is incompatible with the amassing of a large fortune".

But you would not believe this to look at our courts. Indeed they appear to contradict the statement and to show that venality decreases with wealth - or conversely that it increases with deprivation.

So are we there simply to enforce the law? Yes - but with reservations. There are too many laws for any country to enforce them, without much larger criminal justice services, and without the imposition of repression to an extent incompatible with democracy. So law is enforced with discretion - a discretion which is rarely evaluated. Not all laws are enforced and those enforced are not equally applied. That is one problem.

If we shift the emphasis to the courts and declare that we are there to do justice the criterion of evaluation becomes even more imprecise. And if we move on to corrections and ask whether safe custody or reformation is our aim, the objectives get more ambiguous.

The point is that the degree of criminality in a society, the amount of repression, the fear of crime, the sense of security, the extent of justice and the appropriateness of reformation or retribution are all concepts not easily brought within the confines of evaluation. That is not to say that any of them should be discarded, or that ideas of evaluation abandoned. The task is to reconcile them to an extent which can be useful for policy making. For security and justice, rehabilitation or not, the resources are never unlimited.

Whatever our ideals we always have to ask "How much?" - and that implies some form of evaluation, shortcomings or not.

THE MEANING OF EVALUATION

Economics has been defined as the science of allocating scarce resources between alternative uses. It is, in a way, the story of life, because no-one in life has everything he or she wants. Even when everything else is plentiful, time is short and we all have to decide how to make the best of what we have. Inevitably, we - and others - are usually asking how are we doing. Since we are never excused the need to make decisions for our future, we are all consciously or unconsciously evaluating. We do not always do it in money terms but, in a broader sense, we are counting costs and benefits or costs and effectiveness.

This ordinariness behind the complicated techniques of modern evaluation should occasion little surprise. Einstein in his biography defined science as being

> "...The attempt to make the chaotic diversity of our sense-experience correspond to a logically uniform system of thought"

and Thomas Henry Huxley said

"Science is nothing but trained and organised commonsense".

These days evaluation is indeed a science and perhaps something of an art. Evaluation has been defined by the World Health Organisation's Expert Committee on National Health Planning in Developing Countries as

> "...a process which measures the degree to which objectives and targets are fulfilled and the quality of the results obtained. It measures the productivity of available resources in achieving clearly defined objectives".

The word "productivity" here betrays the economic influence : but I have deliberately chosen a health services approach to evaluation to indicate that the meaning of "productivity" must be something a great deal more than "output" in the purely economic sense. Obviously, with health services, efficiency is not easily measured in man-hours of production. Death rates, the distribution of illness and the difficulty of measuring health come into this. Yet, quite obviously, the medical planners feel the need for evaluation despite all the imponderables in which they have to deal. And that is a hopeful sign for those who wish to apply evaluative techniques to criminal justice services. Even if we cannot hope to quantify some of our values, we can hope to develop our own techniques for appraising what we are really doing.

An evaluation is, very briefly, an attempt to judge the effectiveness of a policy, a programme or a single project : but great care is needed in the application of the logic. Let us suppose that an additional one hundred policemen are added to the patrols in a given district and in a period of three months the numbers of burglaries in that area fall. We can, and probably will, jump to the conclusion that the increase in police strength was effective. This will satisfy the undiscriminating and may get votes for politicians or an increased police budget from the Treasury. But a good evaluator will not be satisfied. If he has looked at crime statistics he will be alive to the differences between a snapshort as compared with a moving picture of crime rate curves. We cannot appraise a fall or rise until we set it in a rather longer time perspective. In statistical terms, the three month fall in burglary figures might well be a random variation. And, to understand it, as we have already seen, a time factor has to be considered. Maybe this particular three months fall is a kink in a rising curve from periods of much less burglaries. Then imagining oneself outside the immediate data with which we are presented it is possible to pose some very pertinent questions. Shouldn't more police mean more charges, rather than less. Is it possible to check on the rate at which offences are reported to police. Again, just before

the deployment of the police in the area, was there, by any chance, a professional burglar jailed so that he was inactive during the crucial period. Or, what was happening to burglary in other areas? Had the offenders just transferred their attention to other neighbourhoods? Finally, are we confident that there has been no change in the instrument. A simple legal change could have converted "burglaries" to "housebreaking" or else a shift in the distinction between police reports and crimes could affect the statistics.

Clearly, it will not be possible to answer some of these questions satisfactorily. To obtain such answers to the questions that arise in an evaluation it is necessary for a project or programme to be designed in advance with such questions in mind.

Very simply, a project or programme which we hope to evaluate should be planned in such a way that it will be possible to provide answers to the following questions at specified intervals

1. Was it effective?

   By this we mean, did it achieve its declared goals. Obviously it is desirable to declare the goals in advance if one hopes to evaluate : but there may be objectives

implied in the very nature of a project. It is hardly necessary to stress, for example, the the custodial objective of a maximum security institution. Objectives may be short term or long term and may sometimes be in conflict. For example the welfare of all people is the long term aim of all governments : but often they adopt short term objectives related to economics or defence which seem at variance with the longer term objectives. However, for effectivenss to be ascertained we have to know the objectives, goals, targets. These should be specified in advance.

2. Was it significant?

In other words "so what?" The objectives may have been achieved without there being any significant gains. The successful project may have meant very little to the majority of those involved. Again, hopefully, the likely significance will have received attention during the planning stages, so that in the evaluation, hopes can be compared with achievements.

3. Was it efficient?

Did it cost more than had been anticipated? What were the benefits procured at what prices. Could it have been done differently, more cheaply.

4. Did it comply with the original design?
   Or did it either fail or succeed by departing from the plan and its intended implementation? Did it achieve its objectives by unintended methods? What were they?

5. What were its unanticipated effects? And were there any by products. With so much unforeseen, a great deal can happen which could not be planned but which is significant none the less.

UNRESOLVED QUESTIONS IN CRIMINAL JUSTICE

The most important unanswered question in criminal justice is quite simply "What Works?" Police methods of control and investigation have both remained the same and changed over time without us knowing too much about their effectiveness. After all, is the police objective to be measured by the number of convictions, or by the public sense of security? Do the courts work by hearing cases or by the amount of justice they dispense? Are prisons to be evaluated by the control they exercise over escapes or by the rates of recidivism? What kind of legislation is the most effective? Which kinds of sentences deter? Which punishments reform?

It is instructive to realise that, with all the dignity and majesty of our criminal justice bureaucracies, we still do not know much more about preventing Cain or protecting Abel than we did thousands of years ago. Of course, most of those engaged in the criminal justice services, are convinced of their usefulness. When people are frightened by crime they demand more police protection. Everyone sees the need for a fair trial. Prisons have become more important as society has abandoned the death penalty and corporal punishment. Maybe they don't work : but we don't know what else to do.

The most important central fact is that whatever value society obtains from the criminal justice system, it has little, if any, effect upon crime in the aggregate. We are not even sure of its effect on individuals - good or bad. The evidence is conflicting. The study of 20th century crime in Australia recently completed[1] indicates that if we leave out traffic offences, the rates per 100,000 people of serious violent crime coming before the higher courts have not increased. Just to

---

(1) Satyanshu K. Mukherjee "Crime Trends in Twentieth Century Australia" Australian Institute of Criminology in Association with George Allen and Unwin : Sydney.London.Boston. 1981 p.86

realise that according to victimisation studies our entire paraphernalia of criminal justice deals with less than fifty per cent of many crimes is a reminder that we have expensive machinery which may not be worth what we are paying for it. In Australia since the year 1900 the proportionate cost to the taxpayer of the prisons has risen five times and the cost of the police has risen three times. These are figures corrected for inflation and adjusted for increases in populations and police and prison services.(2) The need for evaluation is therefore very clear. But that clarity does not make it any easier.

We have seen that there are problems about objectives. To evaluate the criminal justice services we have to know what they are intended to do. Maybe they do not greatly effect aggregate crime: but are they intended to do so? Are not the public relations side of police work, the standards set by the courts and the deterrent effect of the prisons reason enough for the existence of our criminal justice services. If so, then the objectives should be clearly stated and evaluations conducted on these lines.

---

(2) Unpublished work by the writer on "Costs of Crime in Australia"

Within this general pattern there are other questions awaiting elucidation by evaluation. For instance, to what extent do the usual management procedures efficiently serve the disciplined forces such as the police and the prisons. For years the upper echelons of the police have been moving from operations into law, science or management. They have specialised as prosecutors, have become experts in finger prints, photography, chemistry or computers or else they have gone into management in a big way. But management in a business aimed a profit or administration in a bureaucracy aimed at dispersing responsibility may be very different to management in a disciplined force. The criteria for success may be different. The procedures may be unique. This is very likely to be the case now that unions are exercising managerial power and influencing policy decisions. Maybe the criteria for successful management may be not only effective operations but maintaining the balance which permits operations to pursue its own objectives. Little work has been done on the effectiveness of career structures within both police and prisons, on the possibility of using external educational qualifications as criteria for promotion, on the effectiveness of merit as opposed to seniority systems. So much has been taken for granted and often there is little room for manoeuvre.

Within operations there are any number of unanswered questions - such as whether it is true that more policemen on patrol reduce crime, whether one man radio cars are as effective as two men cars, whether specialist squads like vice units, homicide bureaux or juvenile services work better than dispersing such responsibilities through a larger general force, whether new computerised systems are as comprehensive and helpful as they appear. We are not even sure what happens when there are no police at all. It is true that in some towns, police on strike have seen outbreaks of looting and theft : but on at least one occasion in New York a police shift not turning up for work seemed to have little effect. The cynic may say of course that that particular city is so far out of control that the presence or absence of police is unlikely to make any difference. But it is important to observe that in general a city goes about its business in a spirit of law and order whether or not the police happen to be in evidence. Indeed the very fact that there is drama when they are called is a testimony to the general orderly tenor of city routine without them.

The U.K. Home Office Research Unit has recently conducted a survey of all available research on police effectiveness in controlling crime and has concluded that the police may be a good deal less effective than is traditionally supposed. It suggests that both community policing and reducing

the public fear of crime may be more important functions for the police.

In rural areas the relationships between police and public are typically so different that questions of comparative efficiencies and methods of measuring them become intriguing. If rural policemen are able to better use the community strengths to contain crime and deny success to the offenders how is this done. Would it happen without the police? If not, what are the methods and what criteria should be used in evaluation. In deciding on promotions and advancement how does the force assess the merits of a police officer. If it is not the number of charges or commendations for efficiency in particular cases and if the examination system does not cover the whole pattern of evaluation, what other qualities are taken into account? Are these appraised subjectively or are there objective measures? Or again, what research has been done on the selection board procedure. Have its successes and mistakes been followed up? If not how can this be done?

We have taken a great deal for granted in the courts but the precise relationship between courts and police has still

to be properly investigated. A great deal of police time is devoted to court procedures. Little work has been done on comparisons of different escort procedures. Are escorts provided by police more efficient than those provided by the prisons? What about the relative effectiveness of bail and remands in custody? For which types of offenders are they best suited? We have some rules of thumb which guide magistrates and judges but a thorough going evaluation of what we are doing now has not been attempted in many places. If the courts are to do justice, what is the effect on this of the adversary system of trials. How many cases have been followed up to check on the justice of the decisions? What evaluations have there been of cross-examinations, of the role of confessions in the total court process, of the value of ritual and formality in court hearings? We assume a great deal but it is paradoxical that a court system which requires everything to be proved beyond a reasonable doubt should have taken for granted so much about its own contribution to justice, peace and order.

Much more attention has been paid, over the years, to the role and purpose of corrections but confusion over the aims of punishment has made conclusions difficult to draw. The obvious preference in the literature for an outright rejection of prisons

as a form of handling crime which does not work, is partly contradicted by the greater use of institutional treatment in the U.S., the U.K. and lately in Australia (though here it is too early to perceive a trend because over previous years imprisonment in Australia has been falling). The U.K. Home Office Research Unit urges the less frequent use of imprisonment and greater efforts to reintegrate those who are imprisoned. It calls for more day centres and shelters to make imprisonment less likely. Social control and welfare considerations compete for attention in the correctional field. Are strict or lenient regimes more effective? What is the precise value of labour, not only for production but for discipline? What are the relationships between prison officers and prisoners? Which ones work better than others. How should prison staffs be trained? What are the advantages and disadvantages for a prison of specialised counselling services, prison chaplains, conjugal visits, home leave, half way houses? The list could be greatly extended.

GOALS AND OBJECTIVES

The critical importance for evaluation of being able to measure performance against clearly defined objectives can

never be over emphasised. In the United States "goals" have been thought to have more generality than "objectives". The California Council on Criminal Justice has offered the following definitions

> "Goal - A statement of broad direction, general purpose or intent. A goal is general and timeless and is not concerned with a particular achievement within a specified time period.(3)

> "Objective - A desired accomplishment which can be measured within a given time frame and under specifiable conditions. The attainment of the objective advances, the system toward a corresponding goal."

Obviously in establishing goals and objectives the particular services will be constrained by the laws which govern it or which establish it as a legal entity. If we hope for integrated planning within criminal justice systems then it would seem appropriate for the goals and objectives of each separate service to be developed in cooperation with - or at least with

---

(3) In the light of the foregoing discussion one could take issue with a "timeless" goal but it is sufficient here to draw attention to the point.

reference to - the goals and objectives of the other services. If police committed to deterrence are seeking maximum arrests (i.e. to fill the prisons) whilst corrections are obliged to release offenders by court order, because of violations of human rights likely to occur from overcrowding, there is a serious problem with a revolving door defeating all attempts to achieve anything but very short term goals and objectives.

Comprehensive goals or objectives for a total service will affect the formulation of goals and objectives for the constituent parts of that agency: but objectives particularly are not merely policy directives issued from the top. Hopefully, in the setting of these general goals, there will have been discussion at all levels : but, once they are set, their achievement depends upon appropriate objectives being decided upon by every department and unit of the particular agency.

Unreality has to be avoided at all costs if any worthwhile evaluation is to be attempted. Idealism is all right, but it has to be given concrete form, in so far as this is possible. This is not to say that goals and objectives have to be chosen because they are easy and readily achievable. To have

an effect on the improvement of a service they may need to offer something well worth striving for. They may need to present a challenge. Correspondingly the goals or objectives should not be fixed so far beyond the possibility of attainment that they become purely academic, unrelated to the daily routine or perhaps likely to engender frustration, disappointment, and discouragement.

Evaluation thrives on quantification. Therefore if at all possible the goals and more particularly the objectives (in the sense in which the term is used here) need to be measurable. Not every goal or objective can be measurable and attention has already been drawn to the change of valuing only what is quantifiable. Yet there should be some way of assessing progress, some way of establishing bench marks from which movement can be measured.

Finally the goals and objectives are not immutable. They are not so sacrosanct that they cannot be changed when necessary to meet altered circumstances. Evaluation is not necessarily deferred by adjustments of this kind. Indeed the decision to make a change is itself a factor in an evaluation.

Obviously budgets require a review of goals and objectives from time to time and the varying availability of resources requires a reappraisal. However, the enormous efforts made to develop programme budgetting with periodic reviews of purposes have largely foundered on the extent to which budgets are already committed to inescapables or have foundered on the descending staff ceilings of governments committed to a reduction in public spending. To some extent these constraints have meant greater interest in getting maximum value for expenditure. This in itself has encouraged reviews of goals and objectives in the light of reduced funding. It has forced the setting of priorities which again helps to make evaluation more precise.

Finding effeciency in a reduced model, however, with more modest goals and objectives should not blind us to the opportunities lost by the necessary reduction of the scale of operations. Here again imagination is needed. Evaluators need to be aware at all times of what the economists call opportunity costs i.e. (in this connection) what we may be losing by small gains.

In setting goals and objectives it is necessary therefore to have reference to the means. The ends and means are necessarily related in any realistic formulation of goals or objectives, targets or aims. Evaluators have to be alive to the way in which one feeds into the other since they are not always separable - and reliability is affected. For example, an objective of the correctional services may be to reduce the prison population by, say, 10 per cent consistent with public security. With facilities for parole it might be possible for the objective to be achieved with relative ease : it will depend upon who is setting policies for the parole board. Or again it is not very realistic for police to talk about reducing the crime rate by 5 per cent if this is likely to be achieved by either greater restraint in police investigative zeal or perhaps by a shift of procedural emphasis from "crime reports" to "incidence reports".

So the evaluator needs to be alive to the ways in which the indicators of the degree of success, or failure, in reaching pre set goals or objectives are obtained. Where do they come from, what precisely do they measure. The question of success or failure can be begged in so many ways. Therefore, though completely neutral or independant measures are not always

available, they should be used whenever possible. When external measures are not available for use, the fact needs to be fully explained in the evaluation and the limitations of the internal measures employed, fully explored.

Though the United States invested a great deal in developing goals and standards for Criminal Justice it is significant that it has provided more detailed information on standards than on goal setting. The National Advisory Commission on Criminal Justice Standards and Goals set up in 1971 produced six reports. The overwhelming majority of the recommendations are concerned with standards. Only a few pages are devoted to the formulation of goals and objectives. This is hardly surprising. To little had been done at that time on evaluation for there to be any close scrutiny of the efforts which had been made to develop goals and objectives.

The idea of evaluation had been drawn from economic planning where a 3 per cent rise in a GNP or an equally precise rise in GDP or perhaps in the value of exports or maybe productivity per head could be readily adopted and progress towards it evaluated. Criminal justice has never been so easy because of the value judgements implied. Actually the economic goals were made to look easier to formulate than they are by the

general acceptance of models which hid some of the difficulties. Garbage, guns and illegal drugs could be regarded as increases in the gross domestic product and listed as achievements when they were probably detrimental to the economy : but such itemisation was masked by aggregate figures which looked more meaningful than they were. It has never been possible to treat crime in quite the same way. As shown, crime rates are largely measures of what the police do. So that, work to achieve a reduction of crime rates needs more precision : it needs to include a presumption (and perhaps an instruction) that recording procedures will not change. Correctional objectives are complicated by society's own indecision about what it wants to do with its offenders. These days, correctional services are expected to maintain standards of human rights. These need to be spelled out in some detail for use as objectives - and when spelled out for achievement they may militate against the public expectation of safe custudy. Security and a high regard for human dignity can and often do conflict.

Justice in the courts may seem to be an unequivocal objective : but what does it mean? Can it be quantified? In some respects it can be quantified - by for example, the cost involved and by the financial significance of the benefits. It

can sometimes be assessed by the length of time required by the courts to process the cases. It has been suggested that it can be quantified by reducing the unevenness in sentencing or by assessing the extent of legal aid available. These are aspects of justice which can be amenable to objectivisation : but obviously they cannot become goals or targets without making some formidable assumptions about the meaning of the judicial process and its relationship to justice.

There is no easy guide, therefore, to the criminal justice service wishing to establish its own goals or objectives amenable to later evaluation. Too little experience has been gained for safegeneralisations. The most that can be offered is a broad guide to goal formulation. As victimisation surveys are more refined, however, as measures of the fear of crime in a community are developed and as guidelines for sentencing take shape the prospects of formulating independent and more precise goals and objectives are likely to be enhanced.

THE TECHNIQUES OF EVALUATION

Knowing the goals or objectives and the means available to achieve the objectives it would seem a relatively simple matter for the evaluator to look at how far the project, or programe has gone and to thereby establish its effectiveness, significance, conformity and by products, expected and unexpected. Actually, it is not nearly so simply. Techniques of evaluation are legion and evaluators have a whole battery of tests and measurements to apply, depending upon the nature of the project, the availability of data, and the specificity of the objectives and means employed to reach them. Moreover, in the field of criminal justice there are usually alternative explanations, either built in or implied.

To take a very simple example, it is desired to eradicate muggings in a particular park. The area is relatively saturated with police, plain clothes detectives are deployed, decoy victims provided and a watch kept on all movement into and out of the area. The result is that for a few weeks there are no muggings in that particular park - and there may be a few

persons arrested for loitering with intent or even for attempting to rob. It is unlikely that there will be a completed offence to bring to court anyway, simply because with so many police about it would not be possible for them to stand by and restrain themselves until the robbery was completed. So the offences will be those of persons caught in the act. It is, of course, possible that those arrested may confess to having committed earlier robberies in the park. Whichever way it goes, the two or three weeks crime free park will have been a success.

However, it is not all that simple for the evaluator. The police preoccupation with the park may not only have frightened off the muggers but also their potential victims. With such intense policing of the area it is likely that ordinary people will avoid the park for a long time, even though it is crime free : so that its "cleaness" may be attributable as much to its lack of people as to the police action. These two will be related so that the correlations do not have unequivocal meanings. The police policy has cleared the crime by the simple device of clearing all the people.

Then again, did the policing of the park lead to a simple displacement of crime? Whilst the police were busy there, were the criminals busy elsewhere - perhaps profiting from the relative absence of patrols in other areas?

There are ways of organising the projects so that some of these possibilities are allowed for. However, this is only one of the difficulties which arise when we try to subject, human or social projects to evaluation.

Cost-benefit and cost-effectiveness analysis are practically self explanatory. Both came from the United States - one from the valuation of water resources projects and the other from the Pentagon. As a method of ensuring the profitability of public investments, the U.S. Flood Control Act of 1936 require that the Corps of Army Engineers should not certify a project under the Act unless it could be shown that its total benefits would exceed its total costs. The procedure looked very convincing both scientifically and in business terms. Undoubtedly there would have been some juggling of anticipated returns to obtain the funds for some of the projects. Calculated

returns can be based on assumptions about how people will act - what crops they will sow and how prices will develop. So the cost-benefit analysis may not always have been as neat as it looked. Nevertheless, it achieved glamour as an institutionalisation of rationality and other governmental agencies adopted the technique.

Cost-effectiveness derived from military evaluations of weapon systems. Clearly, different levels and degrees of weapon effectiveness (i.e. destructiveness) could not be regarded as "benefits" or calculated in straight money terms. But costs could be calculated so that different outcomes (however these were described or calculated) could be compared with cost differentials.

In evaluation the purpose of cost-benefit or cost-effectiveness analysis is to maintain a balance between inputs and outputs. So often an evaluative type of research will demonstrate that better results can be obtained from one, programme rather than another - and therefore recommend it - at least by implication - without an adequate review of alternative

costs. Harry M. Levin gives an example of an educational control group study showing that children drilled by computers scored better marks in subsequent examinations than those not using the computers. The implication was that the educational authority should invest in the computer equipment for all the children to benefit. It was necessary to ask whether the same improvement in examination results could not be produced less expensively. A closer study of the two groups tested, revealed that one small group in the larger control had actually performed in the examinations as well as, and sometimes better than, those drilled on the computers. So although aggregates favoured the computer users, this did not apply to everyone when the results were checked in detail. Those not using the computers who had performed so well had been drilled for an extra half hour each day by the teachers - and they did as well as those on computers. In terms of cost it was cheaper to pay teachers for extra drills than to purchase the computer equipment[4].

---

(4) Henry M. Levin "Cost-Effectiveness Analysis in Evaluation Research" in Elmer L. Struening and Marcia Guttentag

That example can be paralleled with one which frequently arises in the field of criminal justice. Crime may be reduced by a crack-down to deter wrong-doers. But before deciding that this is the cheapest way, it may be necessary to find out whether similar or even better results might not flow from enlisting community aid at less cost. Levin himself analyses carefully, in terms of costs, three programmes tested with a view to reducing recidivism and demonstrates that the one proved most successful by a follow-up study was not the least costly.(5) He does mention, however, the familiar problem in all follow-up studies that we cannot be absolutely sure that substituting the persons under the several programmes the results might not have been different.

Cost-benefit analysis is based on the assumption that benefits or outputs can be expressed in terms of market prices. The values are ordinary monetary values. And where such values are not obviously computable in monetary terms, an attempt may be made to reduce them to quantities like man-hours, which may then be translated into financial terms. It is difficult to

---

(5) Ibid pp.94-98 "Handbook of Evaluation Research" : Sage Publications : Beverly Hills and London : 1975 : Vol.2; pp.5-122

measure security or the fear of crime for example. If a particular programme succeeded in changing attitudes or community outlooks in ways that were desirable, but not measurable, the results could be compared with costs - and this would be a cost-effectiveness analysis - as we had with the better results in the examinations when these were compared with varying costs. Changes in attitudes, however, could result in behavioural changes and these may be amenable to calculation in money terms. Thus, the fear of crime, or a feeling of security induced by a particular programme, may lead to the buying of security equipment or to people moving into or out of a neighbourhood and to property values being affected. Then the analysis is a mixture of cost-benefit and cost-effectiveness analysis.

There are many other approaches to evaluation. Sometimes the physical output can be sufficiently impressive for budget purposes. So when it was desired to evaluate the achievements of the U.S. Peace Corps working abroad in the developing countries, efforts were made to calculate the numbers of children they had taught to read across the world, the number of miles of roads built or the numbers of people they had fed.

This, however, is a bit like crediting the Australian Institute of Criminology with a fall in Australian crime during the past five years, without taking account of the work of all the other services involved - and without perhaps considering changes in economic and social conditions which probably had a much more profound effect on crime rates. Here we might remind ourselves of the need to avoid a fallacious method of logic. Post hoc does not necessarily mean propter hoc. That is to say, just because one thing follows another, it is not necessarily the cause of it and may not even be connected with it. The fact that beef prices rise after kangaroo meat has been found in the shipments may not have anything to do with it. It may be more a consequence of drought reducing the supplies of beef coming forward or the rise in freight rates. Policemen are well aware of the pit-falls of circumstantial evidence and are usually trained to be careful of drawing conclusions unwarrantably.

Of course, examinations are forms of educational evaluation. There are also psychological tests of attitudes which can be used to measure changes which might have been effected by particular courses or advertising schemes. Governments are sensitive to gallop polls which purport to measure public opinion on selected issues. All kinds of research techneques have been applied to assess the effects of TV on behaviour - especially young peoples behaviour.

Dr James Tien of Public Systems Evaluation was sponsored by the U.S. National Institute of Justice to produce a paper called "Towards a Systematic Approach to Evaluation Design". He proposes that the evaluation design and the programme to be evaluated might ideally be built together. He proposes concentrating on three key stages namely "programme characteristics, threats to validity and design elements" decisions in any one of these areas would affect the others - if the elements are considered in the recommended sequence. "Programme characteristics" means defining the rationale of the programme, the nature of its operation and its anticipated objectives. These in turn then help to identify later "threats to validity" i.e. the rival explanations of the observed impacts of a programme that may threaten the assumed causal relationships. Once the threats to validity have been minimised the evaluator is better able to select the design elements according with the already defined objectives. Dr Tien warns that among the traditional measures of imput, process and outcome which are design elements there is sometimes inconsistency. For example, a patrol officer's response time or travel time to crime scene really measures process but this is sometimes counted in the literature as output. It is the apprehension of a suspect

which is the real outcome according to Dr Tien. Outcome or impact measures according to Dr Tien include only those related to attitude, behaviour and crime, since the ultimate aim of any criminal justice related programme is to effect a change in one or more of these three sub groups of outcome measures. Finally he argues for the inclusion of what he calls "systematic" measures. Most evaluations he says "are fairly explicit about input measures, less explicit about process measures and somewhat fragmentary about outcome measures". Virtually none have examined the programmer's impact on other organisations over a long time period, its suitability of adoption by other jurisdictions and possible alternatives. This is what he means by "systematic measures" and he believes these would greatly enhance the quality of evaluation.

From all the above it will be clear that evaluations can get highly technical. Statistical devices, mathematical formulas for developing models and various aspects of systems analysis have been developed to hold some variables constant for testing purposes, to provide matrices of inter-relationships and

to arrive at rational costs where more direct approaches were not possible. It is not proposed to enter into such detail here. Ample material exists in the literature for those who require it : and there is ample scope for experimentation in applying these formulae and models to the special areas of criminal justice.

For our purposes, however, it is very important that the principles be clear - that the undulying assumptions be exposed and clarified and that the glamour of method is not allowed to cloud the essential logic. For the purpose of developing methods it is desirable that evaluators should be familiar with advanced mathematical and statistical methods. But the overall direction of an evaluation requires some knowledge of economic patterns of thought and a capacity for criticising the methods used and developing alternatives where necessary.

CRIMINAL JUSTICE EVALUATIONS IN ACTION

In 1979 the Swedish National Council for Crime Prevention reported on a probation experiment. The probation

office in Sundsvall had been given a substantial increase in resources in 1972 (mainly extra professional staff) in order to see if measurable improvements in work with clients could be achieved. Various cohorts of clients who were handled during the period 1972-1974 were subsequently followed up and the results assessed three years later. Comparisons were made first with the probation results achieved in another similar town Karlstad, as well as certain other districts. The main criteria of effectiveness used were reconviction, a reduction of alcohol abuse and work stability. The results of this project, which could well be described as an evaluation of the changes in quality of probation produced by extra staff, were to the effect that no differences attributable to the extra resources could be found.

A similar project called IMPACT was carried out by the U.K. Home Office Research Unit a few years ago. With a view to discovering whether probation with intensive supervision could become a substitute for imprisonment even for hard core offenders, supervision in selected London probation offices, was increased. More staff were provided, more opportunities were created for prolonged contact with clients and case loads were reduced. After some years it was discovered that the

intensification of supervision had no determinable effect on success rates. This finding coincides with other studies of probation in the United States which have shown that the levels of success as measured by reconviction are unaffected by varying the number of cases supervised by a single probation officer.

Tackling the problem of evaluating sentencing from the viewpoint of the offender the U.S. National Institute of Justice commissioned Alt Associates to survey 559 Federal and State prisons in March 1978 to collect data on the conditions of confinement in American correctional facilities. To construct an index of confinement conditions the researchers studied. (6)

1. Density i.e. the proportion of inmates provided with less than 60 square feet of floor space in their units

---

(6) See Lawrence Greenfield "Prison Environments and Sentencing Reforms" in Research Bulletin for June 1981 of the National Institute of Justice. pp.2-5.

2. Occupancy i.e. the proportion of inmates sharing their confinement unit with one or more persons.

3. Use of Disciplinary Confinement i.e. the proportion of inmates specially housed for disciplinary purposes.

4. Use of Protective Custody i.e. the proportion of inmates specially housed for self-protection and protection from others.

5. House Confined i.e. the number of hours per day spent in a confinement unit.

6. Security Class i.e. the proportion of inmates classified as maximum security.

7. Priority of Service Provision i.e. the proportion of employees categorised as service providers (as opposed to custodial, administrative, clerical or maintenance staff).

8. Expenditure per Inmate i.e. the direct costs per inmate per year

For each of these nine variables, national means and standard deviations were computed. Using these as norms, selected individual states and prison's were assigned scores representing the number of standard deviation units from the mean of each measure. Positive and negative scores indicated whether conditions in a prison or state were better or worse than the computed mean. The total score for each prison or state would thus indicate the sum of the individual scores for each variable and a summary of prison life conditions. Eight states were randomly selected for analysis and the results with their implications were as follows.

Findings

Aggregation of Index Scores Across States, Across Similar Prisons, and Within a State*

	Density/ Occupancy	Level of Deviance	Freedom of Movement	Access to Services	Expen- ditures	Total Index Score
Across States						
N.M.	-2.7	-21.5	-6.2	-1.0	- .4	-31.8
Ore.	-1.3	+ 3.2	-1.3	+3.2	- .3	+ 3.5
N.C.(5)	-2.8	+ 1.1	-3.0	- .3	- .4	- 5.4
Fla. (5)	-2.3	- 1.6	-3.0	+2.6	- .7	- 5.0
Mass.	+4.5	- 1.7	-1.6	- .4	+2.1	+ 2.9
Pa.	+6.0	+ .8	- .1	+4.4	+1.0	+12.1
Minn.	+6.5	- 2.9	- .2	+6.1	+1.5	+11.0
IL	-1.8	-13.3	-4.3	-1.2	0	-20.6
Across Prisons						
N.M.Pen.	-2.7	-21.5	- 4.2	-1.0	- .6	-30.0
Stateville (IL)	-2.5	-29.5	- 9.3	-4.1	- .3	-45.7
Oregon State Pen.	-2.0	- 1.4	+ 1.1	+2.2	+ .1	+ 2.8
N.C. Central	-4.2	+ 4.8	-10.2	+4.3	- .1	- 5.4
Raiford(Fla.)	-3.2	- .6	- 1.5	0	-.8	- 8.1
Walpole(Mass)	0	- 6.9	- 9.0	-1.2	+ .9	-16.2
Gratorford (Pa.)	+8.3	+ 3.6	- .1	+3.0	+ .4	+15.2
Stillwater (Minn.)	+7.8	- 6.3	- 2.3	+6.6	+ .6	+ 6.4
Within a State (Florida)						
Apalachee	-4.2	-2.1	- .8	+7.3	- .7	- .3
Starke	+5.2	-3.8	-9.8	+1.9	- .5	- 7.0
Raiford	-3.2	- .6	-1.5	0	- .8	- 6.1
Sumter	-4.3	-6.5	-4.2	+3.0	- .4	-12.4
Avon Park	-5.2	+3.7	- .4	+2.3	- .7	- .3

* Data Collected on 3/31/78

The index demonstrated enormous variations in the quality of prison life across states, across similar institutions and within states. Total scores for states ranged from a low of -32 to a high of +12. Individual prisons varied from a low of -46 to a high of +15. In a single state, prisons containing similar mixes of offenders (violent, property, and public order) ranged from -12 to near zero (or the national mean).

As noted earlier, it has been argued that equity, commensurability, and certainty are key concerns for sentencing reforms - as they relate primarily to time served. Analysis of the index suggests an additional consideration. If prison environments vary in the severity of confinement conditions, to what extent does confinement in one facility as opposed to another represent an additional or different sanction?

A second benefit of the index is its use for identifying and assisting prisons where environments are deficient. Targeting specific institutions or correctional systems for assistance, intervention, or case study may, in turn, help determine what factors contributed to the deterioration or improvement of the prison environments.

The index scores also could be used as a historical benchmark. As a consequence, similar surveys, in the future, could provide equally systematic profiles, assessing temporal shifts in confinement conditions by using current index baselines. In this way, efforts toward meeting correctional standards and complying with court orders as well as changes in correctional practices could be evaluated for their impact on specific environments.

Most policemen have now been made aware of the Kansas City Police Study undertaken by the U.S. Police Foundation with funding from the Federal Law Enforcement and Assistance Agency. This was an experimental attempt to find out if it were true that saturating an area with police reduced crime. The city was divided into three areas. In one, normal policing patterns were continued, in a second area the police were withdrawn and responded only to emergency calls. In a third area the police patrols were vastly increased so that police were conspicious everywhere. Briefly the results showed no great differences i.e. no appreciable effects on crime rates. The methodology has been challenged, however, and the results therefore questioned. A less ambitious replication of this study was conducted in Melbourne a year or so ago with rather ambiguous results. Therefore, it is hardly possible to be confident that crime control is a simple matter of police manpower.

The National Institute of Justice also used Kansas City for a study of more than 7,000 calls for service made to the police - with a view to examining the hypothesis that rapid police response increases the prospects of making an arrest. This 5 year study analysed the data on calls related to crime and non-crime. It examined the relationship between response time and arrests, as well as witness availability, citizen satisfaction and the frequency of citizen injuries during crime or other incidents. In something like half the calls for "involvement" crimes, victims and witnesses waited about five minutes before reporting to the police and this delay had more effect on the probability of there being an arrest than either the police despatch time or the time taken to travel to the scene of the crime. Actually, on-scene arrests occurred in only 13 percent of the total calls and it emerged, contrary to expectations, that only 3 percent of the arrests could be regarded as a consequence of rapid response. Of the arrests related to the speed of response, more than a half were either burglary and forgery or fraud and embezzlement crimes-in-progress. In some situations, of course, such crimes can be reported whilst they are in progress without alerting the suspect. In these cases it would seem to be a combination of prompt reporting and the rapid arrival of the police which has produced results.

A rather different approach to evaluation is provided by a 1973 report of the U.S. National Advisary Commission on Criminal Justice Standards and Goals. Reproduced here are the details of an attempt to evaluate the effect of street lighting on the incidence of crime.

Section 8. Case History:

The Dallas Inner-City Lighting Program, An Example Of Research-Evaluation Field Experiment

Inner-City Lighting Program: An Analysis

The Dallas Police Department, in an effort to assess the deterrence capability of street lighting on crime, is currently assisting the Street Lighting Section of the Public Utilities Department in its Inner-City Lighting Program. The Department has acted in an advisory capacity supplying the Street Lighting Section with information on which crimes would most likely be affected by added street lighting. Now the Department has been asked to review and analyze the data available relative to street lighting as a crime deterrent. This report presents the analysis of that data.

Problem

Does added street lighting reduce the occurrence of street-related crime?

Hypothesis

Added street lighting reduces the occurrence of street-related crime.

Definitions

Street-Related Crime: any of the following: rape, robbery, assault (all), theft from person, auto accessory theft, breaking and entering a motor vehicle (BEMV), and burglary.

Independent Variable: added street lights (the addition of extra street lights to a given area).

Dependent Variable: the number of street-related crimes in the specified areas.

Experimental Areas: Beats 114, 117, 124, and 132.

Control Area: the remainder of the Central District.

Procedures

1. Select the experimental and control areas (determined by the work schedule of the Inner-City Program completion dates and the availability of comparative data—for this reason, the beats whose additional lighting had been completed in 1971 or 1972 were selected).

2. Select the time periods to be covered (January through May of 1971 and 1972 and August through December 19, 1)

3. Gather the pertinent data.

4. Evaluate the hypothesis in light of the data gathered.

5. Accept or reject the hypothesis.

Inner-City Completion Dates

103–111–112*		7–24–70
Police Beat No. or Boundary Street Area between Hawkins-Gaston-Florence-Liberty		12–5–70
111*		12–5–70
From Pacific to Woodall Rodgers and Central Expressway to Lamar		7–16–71
132		7–16–71
114		9–27–71
117		11–26–71
124		2–15–72
121–122–123	Anticipated	7–1–72
136–137–128	Anticipated	8–1–72
311	Anticipated	9–1–72
332	Anticipated	9–1–72
312, 314	Anticipated	9–15–72
Fair Park Vicinity	Anticipated	9–15–72
113	Anticipated	10–15–72
115	Anticipated	10–15–72
116	Anticipated	10–15–72
118	Anticipated	12–1–72
125	Anticipated	2–1–73
126	Anticipated	2–1–73

*Old Police District Designations

Number of Street-Related Crimes

Month	Beat 114	Beat 117	Beat 124	Beat 132	Avg.	Non-Lighted
Jan. 1971	26	26	29	31	28	21
Feb. 1971	35	46	32	21	34	26
Mar. 1971	28	19	36	23	26	17
Apr. 1971	29	24	13	20	22	18
May 1971	44	40	55	29	42	24
June 1971	49	44	41	30	41	25
July 1971	35	45	57	20	39	26
Aug. 1971	34	39	38	28	35	30
Sept. 1971	32	30	25	19	26	23
Oct. 1971	34	39	46	27	36	26
Nov. 1971	44	29	45	22	35	23
Dec. 1971	40	50	46	18	38	25
Jan. 1972	25	29	52	20	32	24
Feb. 1972	36	37	38	22	33	24
Mar. 1972	36	39	26	26	32	21
Apr. 1972	29	30	34	17	28	21
May 1972	24	40	43	14	30	22

Analysis

Before analyzing the data in the preceding table, some limitations must be kept in mind. First, the quality of the data collected before October 1971 is suspect because prior to that time data collection of the Department's computer files had not yet been stabilized and the data shown in the preceding table comes from those files. Second, due to the staggered completion times of the additional lighting on the experimental beats, not enough crime data is available on the experimental beats for any generalizations to be made from the results of this analysis. That type of analysis should come after all the additional lighting has been implemented and has been in effect for several months (preferably one year). Third, due to the nature of the Inner-City Lighting Program, the careful control on extraneous variables normally exercised in most other experiments was impossible and, therefore, the doubt must always exist as to whether lighting or something else really influenced the reduction of crime in the experimental areas.

With the above limitations in mind, several linear regression analyses were performed on parts of the data in the preceding table to determine how the crime rate behaved before and during the implementation of extra lighting in the experimental areas and what the crime rate was doing in the control area during the same periods. An average value for "street crimes" was determined per month for both the experimental and control areas and these averages are used in the comparisons that follow.

In applying regression analysis to the data, linear regression relationships were computed because only the general directional trend of the monthly values, i.e., the information given by the slopes of the regression lines, is pertinent to the hypothesis being tested in this paper. If the slopes of the regression lines for the data in the experimental and control areas are calculated for the first 5 months of 1972 (the period when extra lighting in all four experimental beats was completed), the trend line for the lighted area (experimental area) has a slope of −0.9 while the slope of the trend line for the nonlighted area (control area) has a slope of −0.7.* Thus, while the slopes indicate that street crimes were decreasing in both areas (the negative sign indicates this), street crime was decreasing at a faster rate in the lighted area.

When the same analysis is applied to the first 5 months of 1971 (before extra lighting was installed), the slope of the trend line for the experimental area has a value of 1.6 while the slope of the trend line for the control area has a value of −0.2. Therefore, during the corresponding period last year, street crime was increasing in the experimental area and decreasing in the control area.

Even when the analysis is applied to August through December of 1971 (the 5-month period immediately preceding the completion of all extra lighting in the four experimental beats), the value for the slope of the trend line in the experimental area is 1.5 while the value of the slope of the trend line in the control area is −1.0. So, during this period too, street crime was increasing in the experimental area and decreasing in the control area.

These results are summarized below:

	Trend Values		
	Jan-May, 1971	Aug-Dec., 1971	Jan-May, 1972
Experimental Area	1.6	1.5	−0.9
Control Area	−0.2	−1.0	−0.7

Conclusion

An examination of the general trends in street crime data during the periods of interest in the experimental and control areas seems to lend support to the hypothesis that extra lighting reduces the occurrence of street-related crime (in the experimental area). However, with the data collected thus far, this cannot be proven conclusively. The data is not strong enough to dismiss the possibility that it is a chance factor, for instance, and not added street lighting that slowed up the occurrence of street crime in the experimental areas (see Attachment). There is enough support of the hypothesis present in the same data, however, to recommend that the lighting schedule be continued and more data gathered to determine conclusively whether or not extra street lighting is an effective deterrent to street-related crime.

*The calculations are shown in the Attachment at the end of this report.

INTERNAL, EXTERNAL, ONGOING AND RETROSPECTIVE EVALUATIONS

As the interest in technical evaluations has developed the methods and types of evaluations have been classified and reclassified reflecting to a great extent the views or vantage points of those who are describing the subject.

Practitioners within a single discipline or single field of service have seen evaluation as a process of assessment within the boundaries of their interests. It is possible to grade projects or programmes according to the lights of experts who have had the requisite training and experience to appreciate the peculiarities and problems of that subject. It would be totally inappropriate to these evaluators to introduce outsiders with criteria and values of their own who may see the work being done in this field in a very different light, and perhaps in an unsympathetic way. For such internal evaluators the justifications for projects and programmes are those which flow from the importance of the projects or programmes to the specialist field - as this might be assessed by an expert who is familiar with this field. A side effect of this single minded approach to evaluation is that the process becomes highly

specific. Usually such evaluators may have little or no claim to a competence in evaluation in other subjects. Sometimes the methodologist become highly specific too. Techniques are developed which may very well be used in other fields but which are not consciously developed as general instruments of assessment.

At the other extreme are to be found the professional evaluators ready to hire themselves to any agency, department or business prepared to trust them to examine the prospects and programmes, to chose what criteria should be applied, and to present the results for the specialists to consider. Here the interest is in procedures and techniques which can be applied either across the board or in a variety of difference specialist situations.

Obviously both external and internal evaluations will be found in general practice. There is scope for the internal exercises designed to test the consistency and validity of programmes or projects within a particular department or agency. But the problem of a judge being a judge in his own cause readily applies when Treasurers or outside funding bodies are asked to support projects or programmes purely on the basis of internal

evaluations. The criteria which the funding bodies have chosen for distributing resources may be quite different to the criteria adequate for an internal appraisal - and which is based on professional values within the field. Therefore, whilst internal evaluations may be useful for the optimal distribution of resources already available within the specialist domains, external evaluations will probably be necessary to justify extra funds from governments or banks. What would satisfy professionals in their own field will not necessarily satisfy those obliged to distribute limited resources between these professionals and professionals in other fields equally convinced of their right to priority attention.

In the general development of evaluation techniques a number of approaches have evolved. Some of the cost-benefit analyses have looked forward, justifying the future financing of projects and programmes by anticipating returns in future years. The investment in the London Underground and the proposals for hydro-electricity schemes in Tasmania have been justified by the calculated returns which outweight costs in the years ahead. Most evaluations have been retrospective - looking back at what has been done and measuring it against the costs involved. But then there have been ongoing evaluations -

attempts to collect the data for an evaluation whilst a project or programme was actually in operation. Sometimes evaluators and operators have worked together. For example, in the Kansas City police response study already described for seven months the researchers collected their data by riding alongside policemen in the cars answering calls, and by noting the presence of witnesses, victims and suspects at the scene of the incident.

There are circumstances in which an evaluation will be a medley of internal and external studies, of anticipatory ongoing and retrospective appraisals. It depends greatly upon the resources and expertise available for evaluation and the importance to the funding agencies of having the projects and programmes thoroughly studied.

As these different approaches are compared with each other and refined in their application to a wide variety of projects and programmes, there emerges a greater appreciation of their strengths and weaknesses both in combination and separately. Different types of expertise develop too, so that

evaluation eventually becomes a team effort. Naturally as evaluation acquires its own professionals it inevitably evolves its own vested interests, internal loyalties and external opposition.

THE POLITICS OF EVALUATION

This exposition of the principles of evaluation began with a warning that evaluation requires a dash of cynicism and a great deal of humility. It is appropriate now to close on the problem of the politics which inevitably infect evaluation as a subject. It would be nice to believe in the utter neutrality of all evaluation and in its unadulterated rationality. It would be reassuring to imagine that policies are formulated on a series of careful and objective evaluations of the experience gained and the options ahead. Alas there is no such perfection or detachment in human or social affairs. Even science favours preconceptions at times.

As indicated already, the very fact that evaluation becomes a distinctive profession gradually generates a form of opposition to it as a method - from those who are conscious of not being included in the judgemental elite. They question the

value of evaluation not only on method but because it really represents a shift of power. No one likes to feel subjected to scrutiny by someone else, however, unbiased or impartial. For those in authority who choose to be guided by evaluations, there is undoubtedly a shift of authority which will emerge more clearly over time. Moreover, it cannot be imagined that evaluations have always been either efficient or effective or that they have been always appeared objectively. Prior decisions to withdraw funds from projects or department's have sometimes been justified by evaluations; professional or even personal jealousies have not always been excluded and inevitably vested interests have sometimes intruded.

Gradually the very professionalism of true evaluation and increasing evidence of its objectivity will give it more claim to the confidence, not only of those who use it, but of those subjected to it. Furthermore, as evaluation's own critics are initiated into the mystique, as they learn more about its principles and procedures - and perhaps as they hire experts to apply the techniques for their own purposes they will be more assured. Nevertheless, when all the careful assessments have

been made, when the greatest care has been exercised to ensure impartiality - and when the validity of the methods used, to help decide priorities or to allocate resources more rationally, is unimpeachable, the subsequent decisions will be made very frequently without any reference to the evaluation. Power struggles will often determine the curtailment or continuation of projects and programmes. Not only management but unions enter at the decisive stages when someone's interest is affected by the rationality. So evaluation, like almost anything else in this world, will never be totally immune from these weaknesses to which humanity is heir - including self-interest, favouritism, preconceived ideas and politics. Perhaps in the long run we should all be grateful that we are not completely enslaved by rationality anyway. It was Oliver Wendell Holmes Junior who said

"Life is painting a picture, not doing a sum".