ACCOUNTING FOR PARTIAL-AND DELAYED-FAILURE IN SOCIAL PROGRAMME EVALUATION

Miron Muskrat*

ABSTRACT

Social programme evaluation is based on well-established principles of experimental, quasi-experimental and pre-experimental design. Professional evaluators are also showing a growing measure of realism in the application of these principles. At the same time, it may be argued that no sufficient account is taken in the context of social programme evaluation of "partial" failure/recovery and "delayed"-failure on the part of service recipients. Studies undertaken in Hong Kong in conjunction with programmes designed to rehabilitate drug addicts indicate that these twin phenomena may be more common than is generally believed. A suggestion is therefore made that they be accorded greater attention in evaluation work.

It has been argued[1] that, for technical reasons; the public budget tends to be smaller than it "should" actually be. This phenomenon is broadly referred to in professional circles sympathetic to the argument as the "bias against public expenditure". The aim of this research note is *to draw attention to yet another kind of bias against public sector activity, viz. the bias against social programmes* (the bulk of these programmes, of course, are concerned with social prevention and rehabilitation). Such a bias is equally prevalent and that the reasons for it are also technical in nature.

---

*Department of Political Science, University of Hongkong.

[1]See: A. Downs, "Why the Government Budget is too Small in a Democracy", *World Politics*, 12 (July, 1960), 541-563.

The bias against social programmes stems from the way in which these programmes tend to be evaluated. The conventional criteria employed in assessing the effectiveness of the various forms of social intervention are somewhat deficient and their utilisation sometimes lead to underestimation of the benefits resulting from allocating resources to improve the social functioning of individuals and groups. These criteria are peculiar to social programme evaluation and the fact that they survived practically intact since the advent of social experimentation is, rather puzzling. This paper advocates the use, whenever appropriate, of additional criteria. First, however, the mechanics of conventional social programme evaluation is briefly described.

Conventional Social Programme Evaluation

Over the past two decades, social problems have become a major concern of public institutions. Alcoholism, crime, drug addiction, education, health (both physical and mental), housing, poverty, social discrimination and a host of other problem areas have occupied headlines and the time of politicians and public administrators. Massive resources have been deployed in attempts to ameliorate all these social concerns. Yet the resources are limited and the problems are large. As a corollary, those who are charged with the responsibility of determining how resources are to be utilised in support of programmes to solve social problems need a methodological framework for ensuring "value for money" in the provision of social services. Social programme evaluation goes some way towards furnishing such a framework.

Social programme evaluation falls into two[2] categories: (1) *outcome evaluations* and (2) *experimental* and *developmental* (E and D) projects.[3] Outcome evaluations seek to measure the

---

[2]Omitted are the categories of *demonstration projects* (which purport to show the administrative and/or political feasibility of potential programmes) and *pilot projects* (in which new programmatic ideas are tried on a larger scale than in experimental and developmental work). These categories are less relevant in the present context.

[3]W. Williams, *Social Policy Research and Analysis.* (New York: Elsevier, 1971). For alternative classifications see, *inter alia*, M. Scriven, "Evaluation Perspectives and Proceedings", in *Evaluation in Education*, ed. by W.J. Pophem (Berkeley: McCutchan, 1974), pp. 3-93. M.E. Borus, *Measuring the Impact of Employment-Related Social Programs* (Kalamazoo: Upjohn Institute, 1979). S. Levitan and G.K. Wurzburg, *Evaluating Social Programs* (Kalamazoo: Upjohn Institute, 1979).

effects of social programmes on their recipients. (Outcome evaluations should be distinguished from a far more common type of evaluation, on-site monitoring, which focuses largely on the use of inputs in service delivery and the quality of service. In the latter case, a monitoring team visits the project, examines as much as possible during a short, thorough inspection, and endeavours in a rather qualitative way to evaluate programme effort and performance. The on-site monitoring evaluation emphasises inputs in asking how well the programme is being administered; the outcome evaluation stresses outputs in asking if the programme changes the situation in a desirable direction). Experimental and developmental projects, on the other hand, aim at establishing the merits of new ideas with programmatic implications in terms of outcomes in a setting corresponding at least in part to actual field operating conditions. An experimental project creates a controlled and simplified field setting in which a limited number of critical variables and the important interrelationships amongst these variables are defined in such a way that the effects and the relationships can be measured in (relatively) precise terms. A Developmental project, in turn, creates a field setting in which the various components of the project correspond to actual operating conditions in such a way that only the total project effect can be assessed.[4]

Outcome evaluations and experimental and developmental projects share a common methodological framework. Under ideal circumstances (Figure 1) this entails following a number of prescribed steps.[5] The first two steps consist of the identification of the target population for which the programme is intended and the drawing of some kind of a probability sample from it to provide individual cases to be used in the evaluation/experiment (the objective of the latter is to ensure external validity for if the subjects

---

[4]The distinction between experimental and developmental projects is pragmatic; ideally one would prefer a real-world setting in which the interrelationships of *all* variables are defined and measured precisely. The complexity, however, of actual operating conditions and the methodological limitations in handling such complexity simply render this "ideal" unattainable.

[5]These steps are described in: T. H. Poister, *Public Program Analysis* (Baltimore: University Park Press, 1978).

participating in the experiment are representative of the target population the findings can be generalised beyond the sometimes narrow domain[6] of the experiment).[7] Given a set of cases to be used in the experiment, the third step is to assign them at random to two groups: an *experimental group* and a *control group* (random assignment lies at th heart of experimental design and its purpose is to eliminate any selection bias).[8] Next, the experimental group is exposed to the programme whereas the control group is not or receives a placebo. Finally, relevant measures are taken after the experiment has been in effect for a sufficient time period for the anticipated results to have occurred. If nothing else is permitted to vary between the two groups, any differences noted can be attributed to the programme treatment (for example, if a probability sample of alcoholics is assigned at random to an experimental group which receives treatment and a control group which does not or receives a placebo, and the rate of recovery is substantially higher[9] in the experimental group, we may infer that the treatment is effective). This experimental design is referred to as *posttest-only control group design* and it can be diagrammed as follows:

	Exposure to Programme	Measurements After
Experimental Group	X	O_1
Control Group		O_2

Occasionally it may be necessary to take measurements before the implementation of the programme (this is accounted for in Figure 1 as well) in which case the experimental design becomes a

---

[6]Which is very narrow in the case of experimental and developmental projects. But even in outcome evaluations no more than a fraction of the target population is usually involved.

[7]The concept of "external validity" is explained in: Poister, *Public Program Analysis*. D. Nachmias, *Public Policy Evaluation* (New York: St. Martin's Press, 1979).

[8]Selection bias threatens the internal validity of an experiment (internal validity refers to the validity of results as they pertain to the specific experimental setting). See: Poister, *Public Program Analysis* and Nachmias, *Public Policy Evaluation*.

[9]In terms of statistically significant differences.

pretest-posttest control group design (or simply the *classical controlled experimental)* and its diagramming assumes a somewhat different form:

	Measurements Before	Exposure to Programme	Measurements After
Experimental Group	0_1	X	0_2
Control Group	0_3		0_4

(Here the effect of the programme is considered to be the difference between any changes that occur in the experimental group $(0_2 - 0_1)$ and any that occur in the control group $(0_4 - 0_3)$. The total effect of the programme, the changes in the experimental group which did not appear in the control group, is therefore $(0_2 - 0_3)$. The total effect of the programme, the changes in the experimental group which did not appear in the control group, is therefore $(0_2 - 0_1) - (0_4 - 0_3)$).
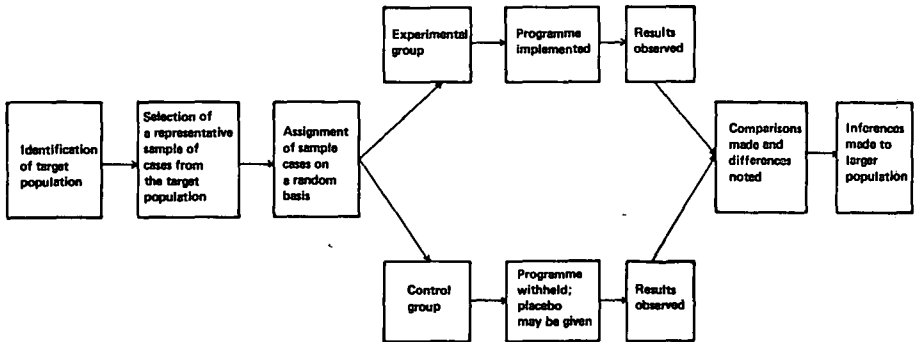

Figure 1

Idealised Experimental Design Process

Experimental designs more sophisticated than the posttest-only group design and the pretest-posttest control group design are also available to the social programme evaluator. The obvious cases in point are the *Solomon Four-Group Design,* the *randomised block design* and the *factorial design.*[10] These designs, however, are generally considered too rigorous and are seldom employed in the context of social programme evaluation (they are deemed more suitable for "pure" social science research). In fact, the trend in social programme evaluation has been towards less rather than more sophisticated experimental designs.[11] For although it is agreed that the posttest-only control group design and the pretest-posttest control group design (the same, of course, applies to the more sophisticated variants) are the optimal means for securing information for making inferences about the effectiveness of social intervention, the practical and ethical constraints facing social programme evaluators are of such magnitude that they are often forced to abandon randomisation procedures and control groups.

The partial abandonment of the full-scale experiment has prompted research methodologists to redefine social programme evaluation in terms of "controlled observation" (rather than "controlled experimentation").[12] As indicated, controlled observation does away with random assignment and control groups. Post-programme performance of the recipients of a social programme is compared to their pre-programme performance[13] (preferably in

---

[10]See: Poister, *Public Programme Analysis.* Nachmias, *Public Policy Evaluation.*

[11]Experimental designs less sophisticated than the posttest-only control group design and pretest-posttest control group designs have been made legitimate by: D.T. Campbell and J.C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1966). See also: T.D. Cook and D.T. Campbell, *Quasi-Experimentation* (Boston: Houghton Mifflin, 1979) and W. J. Duncan, "Quasiexperimental Research in Organizations", *Human Relations,* 34 (November, 1981), 989-1000.

[12]J.A. Caporaso and L.I. Roos, eds., *Quasi-Experimental Approaches* (Evanston: Northwestern University Press, 1973).

[13]Sometimes no direct account is taken of pre-programme performance and post-programme performance is compared to expected performance or a similar benchmark. This can be diagrammed as follows:

$$X \qquad O_1$$

several points in time and not just immediately before and after[14] programme implementation)[15] or to the performance of another group which has not been exposed to the programme but which is not a control group in the strict sense of the word (that is, a *nonequivalent control group*).[16] Such comparisons serve as the basis for making inferences about programme effectiveness. Controlled observation assumes many forms[17] which differ considerably in their sophistication or lack thereof. The more sophisticated types come under the category of *quasi-experiments* and the less sophisticated ones under the category of *pre-experiments* (or *bargain-basement designs*). For reasons given earlier, quasi-and pre-experimental designs

---

[14]That is, after enough time has elapsed for the programme effects, if any, to filter through.

[15]At least three pre-programme and three post-programme measurements are recommended. Or, symbolically:

$$O_1 \quad O_2 \quad O_3 \quad X \quad O_4 \quad O_5 \quad O_6$$

This is the most familiar example of a *time-series design* of which many variants are in existence.

[16]Typical examples of designs involving the use of a nonequivalent control group as a benchmark would be the following:

| Experimental Group | X | O_1 |
| Nonequivalent Control Group | | O_2 |

and:

| Experimental Group | O_1 X O_2 |
| Nonequivalent Control Group | O_3 $\quad$ O_4 |

The former attempts to replicate the posttest-only control group design and the latter the pretest-posttest control group design. The latter can also be extended to form a *multiple time-series design.* For example:

| Experimental Group | O_1 O_2 O_3 X O_4 O_5 O_6 |
| Nonequivalent Control Group | O_7 O_8 O_9 $\quad$ O_{10} O_{11} O_{12} |

[17]Poister, *Public Program Analysis* and Nachmias, *Public Policy Evaluation.*

are more commonly used nowadays in social programme evaluation than their experimental counterparts.[18]

Critique of Conventional Social Programme Evaluation

There exists a decline of experimental rigour and the growing reliance on quasi- and pre-experimental designs in social programme evaluation. This is a healthy as well as an inevitable development and that the intellectual resources should be harnessed to provide support to activities directed at refining the manifold forms of controlled observation rather than to undermine their shaky foundations. The difficulties with conventional social programme evaluation lie, somewhere else and are equally visible in work patterned according to the canons of controlled experimentation. Put differently, the bias against social programmes is not the offshoot of one design or another but can be traced to the whole gamut of evaluation work.

The origins of this bias, are twofold: (1) *the inability to distinguish between the various degrees of failure (that is, "partial" failure vs. "complete" failure) in the performance of the recipients of social programmes* and (2) *the inability to take due account of the time factor in the failure of such persons.* The discipline of quality

---

[18]Many useful examples of social programme evaluation based on experimental, quasi-experimental and pre-experimental work are to be found in: P.H. Rossi and W. Williams, eds., *Evaluating Social Programs* (New York: Seminar Press, 1972).

C.A. Bennett and A.A. Lumsdaine, eds., *Evaluation and Experiment* (New York: Academic Press, 1975).

E.L. Struening and M. Guttentag, eds., *Handbook of Evaluation Research* (Beverly Hills: Sage, 1975).

C.C. Abt, ed., *The Evaluation of Social Programs* (Beverly Hills: Sage, 1976).

F.W. Hoole, *Evaluation Research and Development Activities* (Beverly Hills: Sage, 1978).

J.G. Abert, *Program Evaluation at HEW* (New York: Dekker, 1979).

D. Nachmias, ed., *The Practice of Policy Evaluation* (New York: St. Martin's Press, 1980).

F. Ferber and W.Z. Hirsch, *Social Experimentation and Economic Policy* (Cambridge: Cambridge University Press, 1981).

Another good source of examples are the results of empirical studies reported in article form in the following annual and journals:

Evaluation Studies Review Annual, Educational Evaluation and Policy Analysis, Evaluation and Program Planning, Evaluation and the Health Professions, Evaluation Comment, Evaluation Review, Human Relations, Journal of Social Service Research, Social Work Research and Abstracts and *Studies in Educational Evaluation.*

control (social programme evaluation, after all, is a form of quality control) defines partial failure as "one which does not result in a complete lack of function but which causes the characteristics to deviate from specification limits" and complete failure as "one in which characteristics have deviated outside specification limits and there has been a complete loss of function".[19] Now social programme evaluators tend to mistake partial failure for complete failure and consequently to underestimate somewhat the effectiveness of social intervention. In evaluating programmes concerned with the prevention of drug addiction and the rehabilitation of drug addicts in Hong Kong,[20] there have been encountered several cases in which persons who could justifiably be classified as partial failure/recovery have in fact been classified as complete failure.[21] The effectiveness of the programmes in question, of course, was underestimated accordingly.

The inability to take due account of the time factor in the failure of the recipients of social programmes further reinforces the bias against social intervention. Failure is accorded equal weight by social programme evaluators irrespective of when it actually occurs. This again is contrary to the established principles of quality control which stipulate, as one might exect, that *late* failure is preferable to *early* failure.[22] If we were to apply the same principles in the context of social programme evaluation we would doubtless realise that although social intervention often does not result in a much higher overall prevention or rehabilitation rate than on-intervention

---

[19] A.G. Robertson, *Quality Control and Reliability* (London: Nelson, 1971), p. 140. p. 140.

[20] To acknowledge the assistance of those Master of Public Administration students at the University of Hong Kong who took part on Policy Analysis in 1979-80 and who generated much of the data on which this research note is based.

[21] In studies of human mortality, there is no question of the moment of death of an individual. Similarly, in the types of mortality studies of physical property made in connection with the requirements of depreciation accounting, the moment of retirement of, say, a telephone pole would be evident to anyone who observed the pole being removed from service. Now to determine failure, whether partial or complete, in social programme evaluation is obviously far more difficult. It is by no means impossible.

[22] See, in this connection, E.L. Grant and R.S. Leavenworth, *Statistical Quality Control* (New York: McGraw-Hill, 1980).

(that is, that the failure rate in the experimental group is, in the final analysis, not substantially lower than that in the control group or the nonequivalent control group), it nonetheless tends to delay failure for surprisingly long periods of time. In other words, one of the tangible benefits of exposure to a social programme may be "delayed"-failure, a factor which definitely ought to be taken into consideration in any serious evaluation work.

The experience with programmes designed to rehabilitate drug addicts in Hong Kong is a case in point.[23] The bulk of these programmes might be thought of as consisting of four stages: (1) induction, (2) maintenance, (3) growth and (4) voluntary withdrawal.[24] Each participant could be classified at any given time into one of these stages. Objectives during induction might include regular attendance and a steady reduction in the incidence of urine test results indicating use of illicit drugs. The maintenance objective might be continued programme participation, the absence of positive urine tests and stable social adjustment. The growth stage, by contrast, would be focused on specific improvements in vocational and interpersonal adjustment. Finally, during withdrawal programme staff would help participants maintain a stable social adjustment and continued abstinence from illicit drugs (in the case of the increasingly popular methadone maintenance treatment this takes place while achieving a decreasing dose of the maintenance drug). The above outcomes are routinely monitored in most programmes of that nature and can form the basis for evaluation of their effectiveness.

Now since some of these measures could be obtained from drug addicts who did not take part in any of the programmes at issue, the

---

[23]C.M. Harris and S. Stollmack, "Failure-Rate Analysis in Correctional Systems", in *Operations Research in Law Enforcement, Justice, and Societal Security*, ed. by S. H. Brownstein and M. Komrass (Lexington: Heath, 1976), pp. 143-153.

S. Stollmack and C.M. Harris, "Failure-Rate Analysis Applied to Recidivism Data", *Operations Research*, 22 (November-December, 1974), 1192-1205. C.M. Harris and S.D. Moitra, "On the Transfer of Some OR/MS Technology to Criminal Justice", *Interfaces*, 9 (November, 1978), 78-86. H.S. Bloom, "Evaluating Human Service and Correctional Programs by Modeling the Timing of Recidivism", *Sociological Methods and Research*, 8 (November, 1979), 179-208.

[24]See in this connection: C.C. Attkisson *et al.*, *Evaluation of Human Services Programs* (New York: Academic Press, 1978).

paper employed a variant of the nonequivalent control group design with a view to determining the relative effectiveness of those forms of intervention for which the requisite data were available. Initially, the findings concluded that in many cases, at least in terms of abstinence from illicit drugs, treatment proved ineffective for participants tended to experience a relapse and their long-term rate of failure was only marginally lower than that of non-participants. Yet, further manipulation of the data (by means of failure-rate analysis as suggested by Grant and Leavenworth)[25] threw a rather different light on the effectiveness of practically all of the programmes which I were evaluated. It became apparent that participants more often than not abstained from illicit drugs for reasonably long periods of time − long enough to warrant continuing funding (some programmes, of course, turned out to be more effective in this respect than others). It is conceivable that the delayed-failure phenomenon is common to all forms of social intervention and that social programme evaluators simply cannot keep on leaving it out of their analysis.

Summary

In recent years, there has been a discernible trend towards the utilisation of experimental, quasi-experimental and pre-experimental methods in the evaluation of social programmes. Gone are the days when the main input to programme planning and control in the social services was the general insight of the administrator. This, to all appearances, is both understandable and desirable. The systematic evaluation of social programmes, however, has been generally based on criteria of success and failure which leave something to be desired. Evaluation work may be growing in sophistication but the distinction between partial failure and complete failure remains as rare as ever and the delayed-failure effect is not accorded due recognition. As a corollary, the effectiveness of social programmes is consistently underestimated. It is not too much to ask that this methodological

---

[25]Grant and Leavenworth, *Statistical Quality Control.*

defect be corrected and that the bias against social intervention be
thus eliminated.[26]

26Another methodological source of bias against social programmes may be cost-
utility analysis. See in this connection C.B. Buxbaum, "Cost-Benefit Analysis", *Social
Service Review,* 55 (September, 1981), 453-471. It is interesting to note that methodo-
logical problems encountered in evaluation work are causing a shift towards more quali-
tative approaches. See for example: E.G. Guba, *Toward a Methodology of Naturalistic
Inquiry in Educational Evaluation* (Los Angeles: University of California Centre for the
Study of Evaluation, 1978). G. Willis, ed., *Qualitative Evaluation Methods* (Berkeley:
McCutchan, 1978). M.Q. Patton, *Utilization-Focused Evaluation* (Beverly Hills: Sage,
1979). M.Q. Patton, *Qualitative Evaluation Methods* (Beverly Hills: Sage, 1980). See also:
S. Maynard-Moody, "Reconsidering Charity", *Administration and Society,* 13 (February,
1982), 379-403.

SELECT BIBLIOGRAPHY

Abert, J.G. and Kamrass M. *Social Experiments and Social Program Evaluation.* Cambridge, Ma.: Ballinger, 1974.

Berstein, I.N. *Validity Issues in Evaluative Research.* Beverly Hills: Sage, 1979.

Caro, F.G., ed. *Readings in Evaluation Research.* New York: Russell Sage Foundation, 1971.

Cronbach, L. J., *et al. Toward Reform of Program Evaluation.* San Francisco: Jossey-Bass, 1980.

Datta, L.E., and Perloff, R., eds. *Improving Evaluation.* Beverly Hills: Sage, 1979.

Epstein, I., and Tripodi, T. *Research Techniques for Program Planning, Monitoring and Evaluation.* New York: Columbia University Press, 1977.

Fairweather, G.W. *Methods of Experimental Innovation.* New York: Wiley, 1967.

Fairweather, G.W., and Torantzky, L.G. *Experimental Methods for Social Research.* Oxford: Pergamon, 1977.

Fink, A., and Kosecoff, J. *Evaluation Primer.* Beverly Hills: Sage, 1980.

Franklin, J. L., and Trasher, J. H. *An Introduction to Program Evaluation.* New York: Wiley, 1976.

Freeman, H.E., and Sherwood, C.C. *Social Research and Social Policy.* Englewood Cliffs: Prentice-Hall, 1970.

Hatry, H.P., *et al. Practical Program Evaluation for State and Local Government Officials.* Washington, D.C.: Urban Institute, 1973.

House, E.R. *The Logic of Evaluative Argument.* Los Angeles: University of California Center for Study of Evaluation, 1977.

House, E.R. *Evaluating with Validity.* Beverly Hills: Sage, 1980.

Lyons, G.M., ed. *Social Research and Public Policies.* Hanover: Public Affairs Research Center, Dartmouth College, 1975.

Mayer, R.R., and Greenwood, E. *The Design of Social Policy Research.* Englewood Cliffs: Prentice-Hall, 1980.

Morrall, J.A. *Program Evaluation in Social Research.* Oxford: Pergamon, 1979.

Moursund, J.P. *Evaluation.* Belmont: Wadsworth, 1973.

Perloff, R., ed. *Evaluator Interventions.* Beverly Hills: Sage, 1979.

Posavac, E.J., and Carey, R.G. *Program Evaluation.* Englewood Cliffs: Prentice-Hall, 1980.

Riecken, H.W., and Boruch, R.F., eds. *Social Experimentation.* New York: Academic Press, 1974.

Rivlin, A.M. *Systematic Thinking for Social Action.* Washington: Brookings Institution, 1971.

Rossi, P.H., *et al. Evaluation.* Beverly Hills: Sage, 1979.

Rothman, J. *Social R & D.* Englewood Cliffs: Prentice-Hall, 1980.

Rutman, L., ed. *Evaluation Research Methods.* Beverly Hills: Sage, 1977.

Rutman, L. *Planning Useful Evaluations* Beverly Hills: Sage, 1980.

Smith N. L., ed. *New Techniques for Evaluation.* Beverly Hills: Sage, 1981.

Suchman, E.A. *Evaluative Research.* New York: Russell Sage Foundation, 1977.

Sze, W.C., and Hopps, J.G. *Evaluation and Accountability in Human Service Programs.* Cambridge, Ma.: Schenkman, 1974.

Thompson, M. *Evaluation for Decision in Social Programmes.* Westmead House, 1975.

Tripodi, T., Fellin, P., and Epstein, I. *Social Program Evaluation.* Itasca: Peacock, 1971.

Tripodi, T., *et al. Differential Social Program Evaluation.* Itasca: Peacock, 1978.

Washington, R.O. *Program Evaluation in the Human Services.* Washington: University Press of America, 1980.

Weiss, C.H., ed. *Evaluating Action Programs.* Boston: Allyn and Bacon, 1972.

Weiss, C.H. *Evaluation Research.* Englewood Cliffs: Prentice-Hall, 1972.

Wholey, J.S. *Evaluation.* Washington, D.C.: Urban Institute, 1979.