Chapter 8

Models and prospects for future evaluation

Evaluation has a clear purpose: 'to provide evidence of the outcomes of programs so planners can make wise decisions about those programs in the future'.[1] Evaluation must always be carried out with the aim of providing decision makers with a rational basis for decisions.

All organisations undertake some appraisal of their activities. For some, it is a kind of subliminal activity—a process that goes on in the minds of some or all people in the organisation. At the other extreme, there is the rigorous scientific exercise which is required to satisfy the usual criteria of scientific experimentation. Between these two extremes are to be found many forms of evaluation activity.

The scientific method has an appeal because of its rigour and precision, but its transfer from the laboratory to the community is not easy. People do not follow physical laws like molecules; nor do they behave like rats in a cage! Moreover, the very length of time and the sustained effort involved in ensuring a correct scientific experimental method give rise to pressures for one-time, ad hoc efforts.

Evaluation of health and welfare activities cannot occur in the same 'scientific' sense as laboratory experiments, but some use can often be made of scientific principles in evaluation, especially if built in when new projects are being planned. For example, the use of the scientific method usually involves comparing a group which receives a program with a similar group which is carefully isolated from the effects of the program. However, it could be unethical and untenable to contemplate that any one section of the community should be denied the benefits of a particular program in order to provide a control group for the testing of results. The Committee therefore recognises that, in health and welfare, the use of the scientific experimental method might often be inconsistent with the principles of natural justice.

Other methods which are sometimes adopted use some of the techniques and rigour of the scientific method. For example, pre-program data may be compared with post-program data over a period to determine changes relative to populations not served by the program. While such methods are not truly scientific, they make some attempt to isolate program factors and may be described as quasi-experimental.

Non-experimental research permits two types of evaluation methodology[2]—comparison of the situation before a program with the situation after, and comparison of planned performance with actual performance. Such methods are easier and cheaper than the scientific methods outlined above. However, often they are not planned in advance. Therefore, appropriate data are not always collected, and changes in program participants caused by factors outside the program are not always adequately allowed for.

Evaluation must be tailored to produce appropriate information for decision making. It could, for example, be inappropriate in terms of resources and costs to undertake a rigorous scientific evaluation of a program costing $5000 which could be evaluated with very simple criteria or with 'soft' data—that is, data that may not satisfy rigorous technical or theoretical criteria. Indeed, some programs are amenable to evaluation only by what may be described as 'soft' methods. It may cost too much to collect some data and in some instances the social theory or statistical techniques necessary for optimal evaluation may not exist.

Further, evaluation may not be a simple 'yes or no' exercise: it may indicate continuation of a program if certain changes are made. Where appropriate, evaluation should offer implications for future action. This is important not only to ensure that evaluation is relevant to decision making but also to reduce the level of any threat associated with evaluation. Evaluation should be not solely a matter of survival but also a matter of educating staff and improving service provision.

Some evaluation models

A number of models for evaluation have been identified. Some are described below, though discussion is by no means exhaustive. All the models except one deal with outcome data and goal attainment, but each answers a different order of question to assist in the making of a different type of decision.

Each of the models is discussed as it might operate under perfect conditions. However, there are numerous versions of each which allow for particular aspects to be emphasised or to be given less emphasis in order to lessen cost, to take account of particular circumstances or to facilitate decision making.

Evaluation without outcome goals

While evaluation without goals may appear to be a logical impossibility, the phrase does describe a number of assessment methods that are used:

1. A program may be evaluated on efficiency criteria alone, solely to determine whether resources are being utilised in the most efficient, economic manner. Issues emphasised might include the determination of which ball-point pens are the cheapest, what are proper cash flows and whether they are being maintained or what checks are needed to ensure that there is no excess staff, rather than the determination, for example, of whether appropriately skilled workers are answering clients' needs. Such an assessment may, for example, fail to reveal that clients refrain from using the service as it is not appropriate or relevant or accessible.
2. A program may be evaluated purely on cost–benefit criteria.
3. A program may be evaluated in terms of the processes that go on within it: for example, is the counter service adequate and quick, and how long does it take an application for a cash benefit filled in at the counter to get to head office?
4. A program may also be structurally evaluated. For example, are staff appropriately organised, and is the service housed in an appropriate building?

Despite claims by organisations that they have to use these methods because they are not given statements of goals by policymakers, they are in fact using methods that do have unstated goals. These goals are nearly always least-cost or easier-to-do (process) goals, but they do entail the marshalling of organisational activities in a particular direction.

Social experiment

This involves the launching and testing of new programs, and is a forward-looking kind of evaluation. Its function is to inform policy makers of the viability and effectiveness of innovations before much time or money is committed. A social experiment has two basic features:

1. It is a true experiment, with random assignment of participants, with appropriate control conditions, and with before-and-after measurement.

2. The program is explicitly defined and firmly controlled to ensure that the prescribed principles and modes of operation are adhered to.[3]

A less costly variant is the experimental demonstration. In this, no control group is used, but control is exercised over input and process variables and these, together with output variables, are measured. At times, several variants of the same program are run to gauge the effects of variations in treatment.

Social experiments, however, have a number of drawbacks which, while not peculiar to this model, should be noted:

1. They can be costly.
2. Usually, only a few sites can be sampled.
3. Data showing effects on small scattered groups may not indicate outcomes that would result if areas were saturated.
4. Programs may fail, if not well handled, despite the use of correct techniques.
5. There are ethical problems in giving a group benefits for the purposes of a study only to remove them when the study has ended.
6. Long periods can elapse before results are available to decision makers.[4]

Traditional evaluation activity

This is a catch-all term to describe the methods mainly used until now in studying the effects of an ongoing program. It is essentially a before-and-after assessment of the extent to which goals are being achieved. The stimulus applied can often be imprecise and may change according to internal and external opportunities and constraints. However, traditional evaluation activity can provide good estimates of overall effectiveness. This evaluation model is often characterised by being a one-off type evaluation (see Chapter 2).

Four specific types of evaluation activity can be identified here:

1. *Project rating* relies on assessment of the relative effectiveness of different projects within an overall program. Prerequisites are definition and measurement of environmental variables and definition of short-term measures of output.
2. *Overall program impact evaluation* focuses on a program's effectiveness in reaching its objectives. Output variables need to be defined and appropriate measurement techniques established. Groups for comparison should also be delineated.
3. *Relative program effectiveness evaluation* places the emphasis on determining which alternative program techniques, strategies and methodologies are most effective. This type of evaluation depends on prior definition and measurement of the appropriate environmental, input, process and output variables.
4. *Project evaluation* measures managerial and operational efficiency of a project within a program, and depends on the measurement of key output variables, as well as on the use of appropriate comparison groups.[5]

Accountability system

In this model, evaluative data are built into an ongoing information system to meet continuing decision-making needs.[6] Most organisations already collect data, often in very large quantities, about clients, staff, finances and processes. Relatively simple measures of success can be devised and follow-up data collected on a systematic basis.

Accountability can have a number of meanings depending largely on how the data are specified. Data can be specified so as to:

1. designate staff members or units as accountable for their own performance;
2. measure goal achievement, and perhaps constraints on goal achievement;
3. demonstrate a program's accountability to those to whom it is responsible—for example, to the general public, to its own benefactors, to its clients, to other organisations which refer clients, or to several of these groups at once.[7]

If the last meaning of accountability is accepted, adequate mechanisms for the resolution of goal conflict have to be provided, as it is conceivable that the different groups may well demand that the program aim at different goals. Such conflict can be accommodated by the production of multidimensional information reflecting the different goals. Each group would then be able to judge efficacy within its own framework and decision makers would be able to make trade-offs between the differing criteria.

Relevant interest groups identify their criteria and define the necessary outcome data. The accountability system then permits periodic assessment of how well the program is doing from the standpoint of each interest group. Thus the accountability system must ascertain in advance the needs and priorities of intended users and groups to whom a program is held responsible. This allows for selection of the data that will be assessed and utilised.

With this model, care would have to be exercised to ensure that evaluation for its own sake did not assume greater importance than was accorded to the objectives of the activity being evaluated.

Conclusions

None of the models discussed are mutually exclusive. They represent a scale of progressively increasing precision and complexity, and each is a higher development of those below it in the scale. The choice between models is not made simply by selecting the most precise or the most elaborate. The choice will depend on the types of decision that need to be made for the particular evaluation. For example, depending on the requirements that have to be met, a simple model might produce a satisfactory evaluation at lower cost than if an elaborate model were chosen.

There are large numbers of evaluation models and a variety of ways in which each of these can be used. Which model is appropriate and the way it should be used depend on the skills, understanding and competence of the officers involved, on the circumstances in which the evaluation is undertaken and on the type of information required.

Research into the further development and clarification of these models should be encouraged, as they provide the base from which we make decisions about the expenditure of vast amounts of money—decisions that affect the lives of a great many people.

Recommendations

Resource material is needed to aid organisations wishing to carry out evaluation. **The Committee therefore recommends:**

1. **That the Social Welfare Policy Secretariat, either alone or after appropriate consultations, prepare a document, or a number of documents, outlining the methods available to organisations for the evaluation of their activities.**

2. That the Departments of Health and Social Security provide a consultancy service, free of charge, to enable organisations receiving health and welfare grants from the Federal Government to evaluate their own activities.

REFERENCES

1. Carol H. Weiss, 'Alternative Models of Program Evaluation', *Social Work* 19, 1, January 1974, p. 675.
2. Genevieve J. Knezo, Program Evaluation: Emerging Issues of Possible Legislative Concern Relating to the Conduct and Use of Evaluation in the Congress and the Executive Branch (Library of Congress, Congressional Research Service) (Washington, DC, 1974), p. 58.
3. Weiss, p. 677.
4. Weiss, pp. 677–8.
5. Paul F. Gross, Evaluation of the Adequacy of Australian Health and Welfare Services (submission presented to the Committee) (September 1976), Appendix 2, pp. 2–3.
6. Weiss, p. 679.
7. Weiss, pp. 679–80.