

EVALUATION PROCEDURES AND METHODS

5.1 SOME GENERAL APPROACHES

Various approaches to the planning process have already been outlined in Chapter 3. Before proceeding to further discuss the details of evaluation it may be useful to reiterate some points made in that Chapter and to outline the various approaches to evaluation which will arise.

Obviously the complexity of social programming is such that one cannot neatly compartmentalize the different approaches discussed here. Rather the evaluator must have a broad knowledge of the various approaches and tools available to him and be able to make an intelligent selection from his repertoire to meet the needs of any specific situation.

It will not be uncommon for the evaluator to find that planning has proceeded along what we have called either the ad hoc or straight-line approaches to planning. As suggested above, in this situation the evaluator must undertake a cross-sectional examination of the programme. In other words he will look critically at the operation of the programme at a particular point in time even though he will doubtless be using the available evidence on past happenings within the programme to carry out his evaluation.

On the other hand, in a feedback or inter-active planning situation the evaluator is concerned with ensuring that the right data is obtained and recorded so that the on-going programme can be monitored. The evaluator must then at intervals examine the significance of the information yielded by the monitoring process for programme operation.

These constitute the two basic approaches to the evaluative process. The cross-sectional approach is often used because of the lack of continuing monitoring and the lack of appreciation of the need for on-going evaluation. However, even where there is continuous monitoring and evaluation is built-in as part of the total planning process, there may be certain kinds of data best collected at occasional intervals by a cross-sectional approach. In other words, these two extremes of approach are not mutually exclusive and may even be used together.

There may also be questions which cannot be dealt with simply by examining existing programmes. For example, the evaluator may be asked which of two possible ways of providing services would be the most effective. This

kind of question often means setting up genuinely experimental projects so that alternatives can be put into action and tested on a short-term basis.

Similarly the evaluator may be called upon to measure the community need for a particular programme and relate his findings to assessment of the programme itself. Under the discussion of methodology below it will therefore be useful for us to devote some attention to experimental design and to the measurement of need.

5.2 DEFINING THE PROBLEM

The first issue facing any social researcher is to arrive at a satisfactory definition of exactly what it is he proposes to investigate. Problems are often posed by administrators and policy makers in broad and generalized terms which do not necessarily lend themselves to scientific investigation. Thus the first task of the evaluator, just as the social researcher, is often to analyse the problem and to identify those aspects or components which must be investigated.

Similarly social researchers will differ in the extent to which they may emphasize the need to have a fully formulated hypothesis which they can then proceed to test by research. The evaluator will very often find himself in the position where it is premature to construct a rigorous hypothesis and where, therefore, exploratory study must be undertaken to clearly define areas to which high priority must be given in more detailed investigation.

As an example, let us look at the kind of questions which will be involved in evaluating the effectiveness of a short-run training course for youth workers. The original question posed to the evaluator will probably be in such terms as "Is the course effective?"

Part of the analysis which must be undertaken is to examine the precise meaning in this context of the term "effective". It may mean any one or a number of the following:

How far does the course change the behaviour and performance of the trainees in tackling their chosen job?

How far does the course enable the trainees to better communicate with the organization for which they are working?

How far does the course ensure that trainees will remain in their chosen job for a longer period than those who are untrained?

How far does the course result in greater visible effects of the programme in which trainees are working upon the population being served? How closely do these relate to the goals of the programme?

These, of course, are but a few examples of the varying criteria by which effectiveness might be judged. However, there are many other questions which may be investigated among which are:

Are the trainees undertaking the course different from other workers who have not chosen to undertake training?

Is any difference in the trainees really due to the education received on their training course or is it due merely to the fact that they have been singled out for special attention and hence given recognition?

Does the training programme have negative as well as positive effects and, if so, what are these negative effects?

If one looks back to *Figures 3 or 4* in Chapter 3 these will serve to emphasize the extent to which evaluation should be related back strongly to the goals and policies of the programme being evaluated. Again, to take the example of a short-run training course, evaluation should relate to the programme for which workers are being trained and the task which they are being prepared to do. An examination conducted at the end of the training course cannot be seen as evaluation – it is merely a way to give some formal recognition to successful participants in the training experience.

5.3 THE USE OF AVAILABLE INFORMATION

The first task of the evaluator must be to examine existing records and other readily available information. Not only will this provide guidelines for the planning of any more elaborate methods of evaluation but it may in some cases provide sufficient information in itself.

This data may include the records kept by those concerned with the programme being evaluated; other recorded information about the programme; population statistics and other national statistical data; and the results of other research reports or studies which are relevant.

In particular, let us comment on the fact that such a first examination will probably reveal that the records currently being kept by the programme are either inadequate, irrelevant to evaluation, or do not lend themselves to systematic analysis. All too often record keeping in respect to social development programmes is carried out without proper planning as to the purposes and value of the information being recorded. It is a waste of time and resources to record information unless there is a clear purpose for this and established practice for making effective use of the recorded data.

We are aware of one programme where detailed records have been collected and entered into a computerized data file for the last ten years at a relatively high cost. However, this information has never been systematically examined;

there is apparently no plan for doing so; and one can only question the usefulness of further expenditure in supporting such a system. Another example which will be familiar to many readers is the way in which educational programmes focus their collection of data on such matters as attendance records and the results of internal examinations. Both of these are of little significance for real evaluation. Attention should be given to recorded detail about the socio-economic and other background of those participating in education and relating this to the age at which students "drop out" of the educational programme, their destination on leaving and their success in the labour market.

We would also emphasize that keeping records which will be useful for evaluative purposes is very little more difficult than keeping records for any other purpose or in any other way. Naturally the determination of what should be recorded must be related to each specific project but as a general guideline one must look particularly to those kinds of information which will help to measure the extent to which a programme meets its real goals.

Any such records kept over a long period of time should be planned so that those kept in any one year are directly comparable with those of previous years. Whenever a change is made in the system of recording, the pattern of comparability between the old and the new system of recording should be carefully recorded. We will comment a little further on this matter in the next chapter.

Keeping the appropriate records for evaluative purposes achieves nothing in itself. It only provides information which can be used as a tool in the planning and development of the present or a future programme. Just like any other tool it will be of no use unless it is taken up and put to work.

5.4 OBSERVATIONAL METHODS

The value of merely observing the operation of a programme should not be neglected. This may be disregarded because it is too simple and one may find evaluation by observation discounted because it is perceived as unsubstantiated information. However, simple observation of a programme can be valuable if:

The purpose of the observations and the question or questions which are posed to the observer are clearly defined.

The observation is systematically planned according to a particular pattern which will ensure overall coverage.

The results of observation are recorded in a systematic way and made as objective as possible.

Checks and controls are built in to provide for accuracy, validity and reliability.

In meeting these criteria a number of factors need to be considered while planning the observational approach. The units (individuals, groups, etc.) to be observed should be clearly defined and the times at which they are to be observed properly planned. The particular facets of behaviour or activity which are of concern must similarly be established. Standardised systems for recording observations should be developed and carefully used.

A very simple example which is widely used is the counting of traffic by individual observers. During a traffic study, the observers will be located at precise positions and asked to record all traffic movements of a particular character during specified time intervals. Their results will normally be recorded with the aid of mechanical counters. There is no reason why a similar pattern of operation might not be applied to many facets of human activity. Another example which will be familiar to many personnel in the youth work field is the use of systematic observational techniques in evaluation of group discussion. Many standardised forms have been developed for recording the nature and direction of individual contributions to discussion in order to provide the basic data from which the dynamics of discussion can be analysed by those participating.

5.5 SURVEY METHODS

It may often be necessary to make a survey involving asking questions and/or having printed questionnaires completed by individuals participating or who have participated in the programme. On some occasions it will be possible to use standardized tests or measurements (see, for instance, the study by Payne summarized in Part B under BRITAIN). However, the evaluator generally finds that there are no existing standardized tests which fit the requirements of his operation and it will then be necessary to design questionnaires or interview schedules to carry through the evaluation study.

The basic questions in planning a survey are: What information is to be sought? By whom should it be sought? In what way should the information be obtained? From whom can the information best be obtained?

It is not useful here to attempt to answer any of these questions as the answers will be specific to each project undertaken. General guidelines are readily available in the various text books on survey methods which we deal with in Chapter 7.

However, it is useful to point out that there are some kinds of questions which are not very useful ones to ask of respondents. One of these is the hypothetical question which basically asks "What would you do if?" or, alternatively, "Would you use service if this was provided?" These sorts of questions become completely unreliable when the suggested change

in the situation or the new service is one which is completely unfamiliar to the previous experience of the respondent.

Similarly, it is not useful to ask people what their needs are. In the first place there is the limitation that none of us know what we want but can only want those things of which we know. In addition, any answers to such a question will be in terms of what is socially approved, fashionable, or which the respondent thinks the interviewer would most approve of. We will discuss further below the issues involved in trying to measure needs.

Equally, it is not useful to ask directly why people behave in any particular way. Human behaviour is a very complex matter and it is extremely difficult for any of us to really explain the motivation for any particular part of our behaviour. The answers given to the question of "Why did you?" will almost certainly be superficial and of little real value. Experienced social researchers might gain considerable insight into motivation through a complex study but this is beyond either the resources or the purpose of the usual evaluation programme.

Let us also comment briefly on some of the ethical issues involved. Whenever a survey type of investigation is conducted the persons questioned should be told who is carrying out the survey, what its purpose is, and what use will be made of the results. It should also be made clear to them the extent to which any information they give will remain confidential or otherwise. A study which preserves the anonymity and confidentiality of respondents will probably be more successful in obtaining frank answers, particularly where these imply any criticisms of a programme.

It must also be emphasised that every effort must be made to frame questions so that they are completely neutral and do not suggest particular lines of response. Equally it is important that questions are not open to a wide range of different interpretations. As a simple example, the question "Are you interested in politics?" will be of very little use to us unless we know exactly what each respondent means by the word politics. The word can cover an extremely wide range of interests from purely local matters or even international ones. Similarly the word "best" is a very common pitfall. Depending upon its context and upon the feelings of each respondent, it may mean most enjoyable, most educational, most interesting, most efficient, or any one of a number of other things.

5.6 EXPERIMENTAL PROJECTS

The term "experimental project" is an extremely familiar one. A cynical description of what often happens might be something like this:

The people concerned with programme X find that their programme does not appear to be successful and they feel themselves open to criticism for failure. In looking

for ways to overcome this situation they set up a new programme either by slightly changing their existing programme or by copying something they have heard vaguely about as having been tried somewhere else. This new programme is then called experimental partly because the term suggests they are a progressive organization and partly because the new programme might also be a failure.

Although this statement may sound unduly cynical and negative we have seen exactly this approach on many occasions and fear it is all too common. We even notice in one recent report on such a project that the worker responsible stressed very clearly that the project concerned was not merely experimental, but was a repetition of an idea already widely tested, and that rather than having been financed as an "experiment" his work should have been accepted as a normal part of agency practice.

There are well respected traditions in scientific experiment which should be given attention just as much in matters of social organization as they are in the physical scientists' laboratory. We attempt here to summarize these.

Criteria of Experiment

- (i) Any experiment should have a conceptual rationale which relates it to an existing or proposed new theoretical background. Good practice is dependent upon good theory and the two should be integrally related.
- (ii) It should be possible to set down clearly the hypothesis which the experiment is designed to test.
- (iii) The experimenter should make a thorough study of all previous research and theory relating to his proposed experiment and should establish that what he is doing is a genuine experiment in terms of testing a completely new hypothesis or replicating in a new situation an experiment already tested out in another context.
- (iv) From this study of previous research and theory, the experimenter must develop an awareness of any constraints which must be introduced into his experiment in order to avoid the danger of distortion of experimental results.
- (v) Similarly the experimenter must ensure that the methods used are appropriate to the hypothesis being tested.
- (vi) The results of the experiment must be recorded in such a way that the success or failure of the experiment can be measured. It is essential that the record also includes precise details of the methods used.
- (vii) The appropriate controls must be built in to the experimental design so that the experimenter can isolate the effects of the experiment from events which occur as a result of other influences.

On examining a number of recent reports on experimental projects we find that many of these are simply descriptive and impressionistic. Only one has endeavoured to provide some insights into the basis from which the description was written by asking each of those concerned in the experiment

to set down at the beginning a personal statement of their own attitudes and values entering into the experiment. Equally, very few have had any sort of established control against which changes in the young people served by the experiment can be measured.

However, many of those involved in practical matters will argue that insufficient funds or other resources are available to them to set up controls and that they must therefore rely upon impressionistic evidence. It seems to us that this has important implications for those concerned with the funding of experimental programmes. It would be far more useful to set up one experiment with appropriate methodology and control than to set up 25 without. Similarly any authority concerned with the funding of experiment should ensure that the results are fully reported and published in a form which makes them readily available to others concerned.

5.7 THE MEASUREMENT OF NEED

We have referred above to the difficulties in measuring need for services. This is a particularly difficult area of social research, but one which is increasingly being demanded by administrators and others. A recent paper by Bradshaw and at least two others based upon it, indicate some useful directions which might be followed.¹

Briefly Bradshaw suggests that there are four kinds of statement which might be made about need:

- (1) A particular standard for provision of services might be laid down by experts and then one can measure the extent to which any particular community, region or country measures up to this standard. For example, the standard might be laid down that all young people should have a minimum of six years full time education. Against such a standard one can then measure the extent to which any country falls short of achieving this level. In practice the standard setting approach has a number of disadvantages. There is no common agreement between experts about what is an acceptable standard; furthermore, even with such a standard laid down it is in fact only a target at which to aim, and in most cases once the target is approached or attained, it has lost its value.
- (2) The next statement which might be made is concerned with what people say they want. We have already noted that on its own this is unreliable and of limited value, particularly where a specific kind of service is an unfamiliar one. At the same time, one must acknowledge the importance of a clear demand by people for a service.

- (3) There is also the evidence which might be gained from human behaviour. If we find for instance that there is a vocal demand from a particular community for any one service, but that when the service is available it is not used — then there is obviously a need for re-assessment of the situation. Moreover, behaviour often gives us significant clues to underlying needs about which the people concerned will not necessarily be vocal.
- (4) Finally one can make comparative statements about need. For instance, we might compare two regions and find that in one there is a teacher ratio of 1 : 20 while in another the same ratio is 1 : 35. Again it is relatively obvious that the second region is in need when compared to the first. This kind of data is widely and popularly used but we do note the warning that one must not assume that each service is equally relevant in any cultural situation. Comparisons of this kind are really only valid where the relevant features of the two situations being compared are similar.

The above is a very simplistic explanation and needs elaboration for a full understanding but this can be readily gained from any one of the three papers listed below.

NOTES AND REFERENCES

1. Bradshaw, J., The Concept of Social Need, *New Society*, 19 (496): pp. 640—643 (30th March, 1972).
Mercer, D. The Concept of Recreation Need, *J. Leisure Research*, 5: 37—50 (1973).
Hamilton-Smith, E., Issues in the Measurement of Community Need, *Australian Journal of Social Issues* in press, 1974.