

Chapter 1

The nature of evaluation

Think about the last time that you considered the need to make some sort of change. Before choosing a particular course of action, you would have reviewed the available options, or at least the options that you knew about. You would have assessed how well each option might meet your needs, and at what cost. You would then have weighed up the advantages and disadvantages associated with each of the options before making your decision.

The change you selected might have been about some personal matter such as your family finances, or something to do with your children's future. Or it may just as easily have been related to your professional life. You may have been thinking about introducing a new course, or modifying the student registration system, or increasing student retention. Whatever your area of concern, in order to carry out any change, you will have had to work through the process which we call evaluation.

The process of evaluation which we employ to reach a decision as to the way forward is the same regardless of the area of concern or its source or even of its importance. The care we take, the methods we use and the amount of attention we give to the process in those different situations is another matter. In this chapter we will be looking at formal evaluation, considering the purpose of formal evaluation activities in open and distance teaching organizations and examining the different types of approaches to evaluation which are available to us.

16 / Programme Evaluation and Quality

Formal evaluation

Evaluation then is an activity with which everyone is familiar. The question is, how you can best use evaluation with open and distance learning provision. At the informal level, individual members of any institution will be actively engaged in making their own personal evaluations of activities which come within their own areas of responsibility. The problem will be that, as with all other spheres of life, individuals' perceptions will be coloured and distorted by the particular lenses through which *they* see the world. We can only make an evaluation on the basis of the information to which we have access. The conclusions that we reach will be limited by the quality of that information – its comprehensiveness, relevance, up-to-dateness, accuracy.

A more structured approach

One way of looking at the process of evaluation is to view it as a series of different stages. The stages which comprise this cycle are shown in Figure 1.1. It should be emphasized that reality is usually much more untidy and idiosyncratic. Some stages may be omitted, and the sequencing may not always operate as shown. The old joke about deciding what the conclusions will be before carrying out the evaluation does, as is often the case, carry a grain of truth. For example political pressures may result in stage 7 actions being agreed on political grounds before the evaluation findings in stage 6 are available (a frequent habit with government departments).

Evaluation stages

- 1 Identify an area of concern
- 2 Decide whether to proceed
- 3 Investigate identified issues
- 4 Analyse findings
- 5 Interpret findings
- 6 Disseminate findings and recommendations
- 7 Review the response to the findings and recommendations and agree any corrective actions
- 8 Implement agreed actions

Figure 1.1 *The basic stages of evaluation*

The nature of evaluation / 17

Identify an area of concern

This stage can be triggered in a number of different ways. Formal monitoring procedures such as reviews of pass rates, or course registration figures often identify situations which should be giving cause for concern. Informal means such as letters of complaint, or anxieties expressed by staff can lead to the recognition of the existence of possible problem areas. Cost concerns may result in pressure within the organization for the evaluation of a specific project or innovation, such as the use of interactive video for example. Or again there may be an institutional commitment to provide certain data or certain types of evaluation for external auditing, review or grant awarding purposes. If you think of your own institution, you can probably think of just as many if not more instances where the evaluation process has been triggered by external requests for data or because of political pressure than through the process of objective review. The trigger for the evaluation cycle may therefore operate in a variety of ways.

Decide whether to proceed

Not all problems or potential problems which are identified will be seen as having a sufficiently high priority to warrant further investigation. A decision will therefore need to be taken about whether or not to investigate further, or whether to commit resources for a thorough evaluation.

Investigate identified issues

The ways in which issues are investigated should, wherever possible, be determined by the requirements of the problem. For example, the evaluation of an issue such as the quality of guidance to tutors may be usefully approached using a mixture of in-depth discussion to establish the criteria used by the tutors themselves, plus some quantitative feedback to establish the scale of any particular problem areas.

Analyse findings

Whatever the *type* of study devised and carried out for the evaluation, the data collected need to go through some form of analysis stage. The extent and depth of the analysis will depend in part on the technical competence and in part on the specific interests and institutional requirements of those carrying it out. I have known examples where the analysis of course feedback data was limited simply to a one-page summary of students' written comments presented as a report from the teaching team to 'higher authorities'. I have also seen examples where weeks of sophisticated computer analysis were carried out on complex quantitative data in order to help the course team pinpoint the precise sources of students' problems with a course.

18 / Programme Evaluation and Quality

Interpret findings

The more sophisticated and complex the study, the more important is the interpretation phase. The same set of analyses may well be interpreted in very different ways depending on the particular perspective of the interpreter. A high difficulty rating for a course module may be interpreted as evidence that the teaching approach needs further investigation and possibly some revision, or it may be taken as evidence that the students are insufficiently prepared for the course.

Disseminate findings and recommendations

The dissemination phase can be key in determining whether or not the evaluation findings are used. The timing of the dissemination, the target group for the findings, and the perceived relevance of the findings to people's concerns will all need to be taken into account. For example, the importance of variations in student retention rates may be different for those responsible for ensuring the viability of future courses than for administrators responsible for ensuring adequate provision of exam rooms. The same set of information can carry very different messages to different groups. Increased student retention rates may be good news to some staff in an organization, and a mixed blessing to others.

Review findings, agree and implement corrective actions

These final two stages do need to be seen as part of the evaluation process. Evaluation is not an abstract research exercise but an essential tool of good management. In general the methodologies for the design and implementation of evaluation studies are well developed, but the methodologies for enhancing the likelihood of organizational use of evaluation findings is still developing. Hence the importance of recognizing that these two stages must be included in the cycle.

The purpose of evaluation

The aim of evaluation in the case of any organization must be to support that organization in achieving its goals. In other words, to enable it to become a more effective organization within whatever constraints it has to operate. In educational organizations, the need for formal evaluation activities is usually clearly recognized. In their 1977 review of major evaluation studies, Guttentag and Saar drew attention to the fact that 'education is one of the most highly researched evaluation fields' (Guttentag and Saar 1977).

The learning organization

Evaluation is used, or should be used, to enable institutions to operate as learning organizations. The importance of the role of the detection and correction of error is the basis for the ideas on organizational learning put forward by Argyris and Schön (1978). An important feature of their argument is the view of the organization as a unit or a whole in respect of the reviews of performance and the implementation of subsequent modifications.

For example, individuals or small groups such as course teams may have learnt that the submission rates on assignments for a particular course drop sharply at a certain point. There are a number of possible explanations for this phenomenon which would have to be investigated. It may be to do with the difficulty of the assignment or the course workload at that point. If that is the case, then the person responsible for the course will probably attempt to deal with the problem by changing the assignment or by cutting out some of the student study tasks. However, there may be institutional-level implications for this state of affairs. For example, the number of assignments which students are expected to complete, the monitoring of standards, the course approval strategy and the course testing strategies are all aspects where the institutional procedures may have to be modified if the problem is found to be sufficiently widespread or severe.

Programme evaluation

Programme evaluation in the field of open and distance teaching is relatively underdeveloped. By programme evaluation I mean evaluation which focuses on programmes of study. It is at this level that the pedagogic, management and often the financial responsibilities lie in education and training. It is usually here that responsibility for the detailed issues of quality and accountability have to be exercised.

I have chosen the term 'programme of study' to describe sets or groupings of courses. Usually, these would be sets of courses which share some sort of common aim. That aim may be the award of a qualification for students who successfully complete a requisite number or series of courses in an area of expertise; or it may be that a particular audience is targeted, or a particular teaching medium is used.

Within any institution it would be a simple if onerous task to list large numbers of possible issues to which evaluation could make some contribution. However 'busyness' is no substitute for purposeful intervention at key points. The question then is how to determine what the key

20 / Programme Evaluation and Quality

points are — how are we to identify the purposes of evaluation in such a way as to achieve the best match with the goals of the institution?

Diverse institutional goals

The overarching aims of a provider of education will be related to the provision of learning opportunities and to such associated activities as the accreditation of learning. But such global aims can also contain a diverse range of subsidiary goals. In an earlier work I discussed the different types of goals that learning providers can hold (Calder 1993). Four distinct groupings can be identified:

- society/economy centred
- institution centred
- subject centred
- learner centred.

The society/economy centred goals refers to the skill centred education and training which both public and commercial providers are increasingly encouraged to offer. Institutional goals can include institutional survival; high status among clients, other providers or funders; or public recognition. Providers may also hold 'subject centred' goals, by which I mean claims to scholarship and the desire to provide courses of a high academic quality. The learner centred goals emphasize the personal development aspect of learning and the need for learners to achieve not only subject knowledge and skills but also more sophisticated learning strategies and such intangible outcomes as self confidence, recognition of self worth, and a commitment to the community.

You may have noted the absence of student performance from the list. In the UK, the assessment of student performance is referred to by the term 'assessment'. The term 'evaluation' refers primarily to the evaluation of the teaching and organization activities which support student learning and includes the assessment of student performance as just one aspect or function. However in his book on the assessment of students, Rowntree highlights the fact that assessment and evaluation are often treated as 'virtual synonyms'. As he points out, there are many countries, including the USA, where the term 'evaluation' is used to describe both the assessment of individual student performance in terms of what they have learnt or accomplished and the evaluation of the teaching and other organizational activities which support student learning (Rowntree 1977). In fact some institutions use the term 'evaluation' solely to describe the assessment of student performance.

Needless to say, such differences in the the way the term is used can on

occasion lead to considerable confusion. Discussions about 'evaluation' between professionals from countries separated by different traditions of usage of the same term can be enlivened by the misunderstandings caused by failure to check on the definitions of apparently common terminology. In this book, I will stay with the UK meaning of evaluation.

Diverse interest groups

The particular interest group which sponsors the evaluation is of particular importance in determining the purpose of any particular evaluation activity. Kogan (1989) described well the complexity of the way in which the nature of the evaluation is determined when he commented that

The nature of the evaluation will vary according to whether an intervention is primarily directed to, for example, improvements in quality, reduction in cost, equalisation of access, or improvements in working conditions; and it will also vary according to its sponsors whether they be managers, political leaders, client groups, or the workers who are subject to the evaluation.

What Kogan was drawing attention to was that evaluation is not a clear-cut straightforward activity. Rather the primary purpose of the evaluation and the particular interests of the sponsoring group initiating or sanctioning the evaluation will combine to define what kind of approach, what kind of focus the evaluation will have.

Even where the evaluation is commissioned internally and carried out internally, there may still be great differences in its nature. Consider for example, a situation where the quality of the teaching received by students is being evaluated. If the aim of the evaluation is to assist with staff development, then its nature will be rather different than if its aim was to collect data to use for staff appraisals. This particular example is an important one because many staff have relatively little experience with open and distance teaching, and are frequently unaware of the rather different needs of home-based or 'distance' students from those of conventional students or trainees. Certainly the system of student feedback on teaching used by many providers for assessing face-to-teaching can conflict with the need to use evaluation for staff development purposes.

Approaches to evaluation

We have discussed the fact that evaluation is a process which can be utilized across the whole range of activities in an educational institution. The

The nature of evaluation / 13

Evaluation Purpose	Materials Development Phase	Materials Presentation Phase
Formative	Developmental testing	Rolling remake
Summative	Market testing	Validation review

Figure 1.2 *Examples of activities with different evaluation purposes at different materials production phases*

It would be a mistake, however, to think of the distinction between the two forms of evaluation as formative if carried out during the development phase and summative if carried out during the presentation phase. Figure 1.2 shows examples of both formative and summative approaches being used during materials development and materials presentation stages.

In a similar way, material which is gathered for formative purposes may be used for summative decisions, just as data which is gathered for summative purposes can be and often is used in formative ways. Tessmer, in his book on formative evaluation, points out how 'As long as the purpose of the evaluation is to "revise" the instruction by reorganizing or supplanting it, the evaluation can be a type of formative evaluation' (Tessmer 1993). He gives the example of instructors who may wish to evaluate a 'bought in' course. If they intend to modify or supplement those parts which they consider inadequate for their learners, then *they* would be carrying out a formative evaluation. There is also the point, however, that if it fell below the expected standard, whether technically, pedagogically or academically, then it might be expected that the instructor would decide not to use it at all. In other words, the evaluation would be transformed into a summative evaluation.

Context, input, process *and product*

Pretest — posttest approach

The next step to consider is what methods of enquiry you can actually use in carrying out evaluations. There is a long tradition of trying to set up experimental designs, or the nearest thing to them that was actually feasible,

24 / Programme Evaluation and Quality

in education and media research. However this approach does have limitations because of the problems of trying to control all the variables except for the experimental one. It is also open to criticism about the appropriateness of the 'lab-based' approach for investigating the effectiveness of instructional materials used by different kinds of people in different ways and in different settings. However, variations of it are still used for the formative development of instructional materials. Barbara Flagg (1990) describes a typical pretest-posttest investigation which was designed to look objectively and in detail at what pupils had learnt as a result of the use of videotapes and print materials:

An example of the one-group pretest-posttest design is the formative evaluation of Systems Impact's prototype videodisc lessons on fractions. . . Teachers presented a series of daily lessons on fractions using videotapes and print materials to mimic the instructional design of the Level 1 videodisc.

Criterion-referenced tests¹ integrated into every fifth lesson and comprehensive pre- and posttests established the degree of mastery of the fraction concepts. These tests gave evidence as to what programme content was or was not being successfully communicated.

Figure 1.3 illustrates this approach. As you can see, the learning experience, together with any other events or processes which might take place between the pretest and the posttest are not taken account of. In effect, the interaction of the students with the programme is treated as if it were a



Figure 1.3 *The pretest-posttest approach*

This approach does have a number of methodological drawbacks. Flagg describes problems such as the drop-out from the test group, possible effects of external events, such as TV maths programmes at home, or extra help from parents, and the effect on the group of constant testing. As Flagg points out 'The pretest-posttest objectives-based study has limitations, . . . in its utility for formative evaluation because it provides little insight as to *why* the programme might be working or might not be working.'

¹ Criterion-referenced measures assess a student's achievement of subject matter or a student's behaviours in relation to a criterion standard of performance, not in relation to the performance of other students on the same test.

The nature of evaluation / 25

Illuminative evaluation

Concerns about methodological problems and the recognition of the importance of understanding more about the process which the learner was actually going through led to the development of a very different methodological approach, namely illuminative evaluation. Parlett and Hamilton (1972) who developed and introduced this approach saw the pretest-posttest approach as 'a paradigm for plants, not people'. They wrote:

such evaluations are inadequate for elucidating the complex problem . areas they confront and as a result provide little effective input to the decision-making process.

Illuminative evaluation is introduced as belonging to a contrasting ' anthropological' research paradigm. Attempted measurement of ' educational products' is abandoned for intensive study of the programme as a whole: its rationale and evolution, its operations, achievements, and difficulties. The innovation is not examined in isolation but in the school context or 'learning milieu'.

They explain:

Observation, interviews with participants (students, instructors, administrators and others), questionnaires, and analysis of documents and background information are all combined to help 'illuminate' problems, issues, and significant program features.

What Parlett and Hamilton were identifying was the importance of the process as well as the input and the outcome. There is also a recognition of the importance of the context in which the learning occurs. Figure 1.4 illustrates the illuminative approach.

Clearly there are limits within programme evaluation as to how much of the programme as a whole can or should be evaluated over extended periods. The illuminative approach was developed very much as a response to the 'agricultural-botanical' approach which had previously predominated. The concern with description and interpretation rather than measurement and prediction, however, reflected a substantial shift in evaluators' understanding of the potential of formal evaluation as an aid to decision-making through greater understanding of what happened within educational programmes.

The CIPP approach

For the evaluation of some projects, an evaluation of the context in which it is operating is essential. A major evaluation of the use of the Canadian Hermes satellite for educational purposes in the late seventies drew

26 / Programme Evaluation and Quality

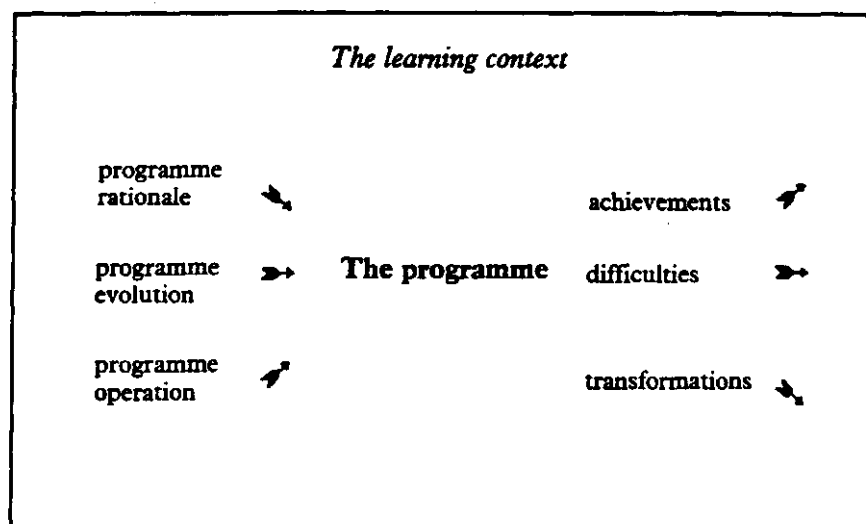


Figure 1.4 Illuminative evaluation

particular attention to this aspect. As the **evaluators explained**:

The importance of context in **educational** satellite projects is **such that** all **evaluations** must be **partly** illuminative, even if **they** rely heavily **on** **survey** methods. The study **of** context **involves looking** at the costs of the project **and** the manner in which it was **managed**. **Although such** issues **are fraught** with **controversy**, **knowledge** of them is essential **to** a fair assessment of **an** experiment. (**Richmond and Daniel 1979**)

The **evaluation framework** which was **chosen for** this massive project was the CIPP **approach** put together by **Stufflebeam and** his colleagues (Stufflebeam et al 1971). **Richmond and Daniel explain** how this acronym describes the **four evaluation** stages which **can encompass** the main aspects of the presentation of **a course, programme of studies or** major project.

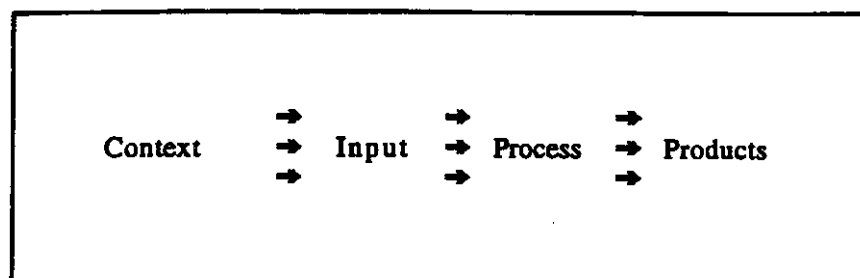
Context evaluation: Descriptive **data about** the **programme** objectives, intended outcomes, criterion measures.

Input evaluation: The selected **programme** strategy.

Process **evaluation:** The **implementation** of the **programme** **procedures and strategies**.

Product evaluation: The success of the **programme**.
(The same as summative **evaluation**.)

The nature of evaluation / 27

Figure 1.5 *The CIPP evaluation model*

The advantage of the CIPP approach is its comprehensiveness. The example given is of a major national multi-programme project, but the approach is one which can be usefully drawn on even for small studies such as a course or even a module evaluation.

Other approaches

We have looked at the main key evaluation approaches, but in practice there are a whole range of different approaches whose usefulness will depend on the particular concerns you have and on which you want the evaluation to shed light. New approaches to evaluation continue to be developed. With open and distance teaching, the role of evaluation as communication between organization, students and tutors becomes more important. Similarly, developments in research methodology also open up new ways of looking at evaluation. For example, the developments in participative research give respondents more power and a greater say in the research process. In the evaluation context this approach can be seen as a development of the illuminative process. We will be looking at some examples of this approach in Chapter 7.

Utilization of evaluation

The final and often neglected phase of evaluation is its utilization. Within institutional evaluation especially, the view of evaluation as an activity which is completed when the final presentation is made, or the report is presented, is too limiting. The circle must be squared by looking at utilization as part of the evaluation process. This final phase is essential if the institution wishes to identify itself as a 'learning organization'.

The need to take seriously the issue of when and how evaluation findings are used by the organization is highlighted by the experience of those

28 / Programme Evaluation and Quality

innumerable organizations who have participated in evaluations of various kinds. A report from the Centre for Higher Education Policy Studies (CHEPS) in the Netherlands made the distinction between three types of results arising out of the evaluation activities associated with their quality assessment system: *no utilization*, *passive utilization* and *active utilization* (Westerheijden, Weusthof and Fredericks 1992).

No utilization

'No utilization' describes the situation where the organization takes no account whatever of the evaluation findings. This situation is, unfortunately, not unusual. Many readers will, I am sure, have shared my experience of having seen decisions taken just before the completion of a major, carefully designed and carried out evaluation study! In such situations, the evaluation report may not even be formally considered within the organization, but merely be put on a shelf; or lodged in a library.

Passive utilization

'Passive utilization' describes the situation where the evaluation findings are formally received and discussed within the institution, without any actions being taken to change anything directly as a result of the evaluation study. The report may be formally disseminated, may be discussed in committee, and may even be the basis for recommendations for a variety of future changes.

Active utilization

Only activity which takes place as a direct result of evaluation findings can be classified as 'active utilization'. Thus if a course evaluation suggests that a particular part of the course is presenting students with difficulties, and as a result, that course is modified, then this would be seen as 'active utilization'.

As Westerheijden et al suggest, this classification refers to the short-term use of evaluation. Very often the relevance and the implications of evaluation findings are not taken up institutionally for some time, but nevertheless, the information and insights from the evaluation may affect thinking within the organization to a considerable degree. You may be able to think of some instances where you have had this experience yourself

While the worst instances of the lack of utilization of evaluation findings are often found in studies which are carried out by individuals or groups external to the institution this need not always be the case. Utilization can fail to take place regardless of whether the sponsoring group for an evaluation activity is internal or external to the organization; the evaluation activities are carried out by individuals or groups who are internal or external to the organization.

The need for agreed procedures for dealing with the conclusions and recommendations from evaluation studies which are linked with the decision-making structure of the organization is clear. Otherwise, whoever the sponsors are and whoever the evaluators are, it can be too easy for the findings of the evaluation to be set aside.

Conclusion

These developments are part of the growing recognition that no stage of the collection, analysis or utilization of data is a value-free activity. The decision to collect information about, say, the age of students, means that the institution considers this data relevant and important. Equally, the decision not to collect data — about dependent relatives, for example — means that this information is not seen to be of importance to the evaluators (although it may be of considerable importance to students who have to make arrangements to get to study centres or to attend residential schools). Such data may be used either formatively, in that the institution may use them in planning or designing provision which is more appropriate to students needs, or summatively, as in monitoring whether certain agreed admission targets have been met.

We must therefore come to the conclusion that not only does 'evaluation' mean different things to different people, but that its definition depends on one's philosophy of education and on how one intends to use the acquired evaluation information. The information which is available about evaluation approaches reflects very much, as might be expected, the predominant concerns and cultural values of the time. So in education, for example, it could be argued that concerns about the quality of the learning experience for students have, to a considerable extent, been superseded by concerns about the efficiency of the providers.

Much of the early evaluation work in the field of education and training was concerned with judging the outcomes from innovative experimental projects. Kogan (1989) refers to the 'massive American literature concerned with the evaluation of large-scale experiments which are undertaken under controlled conditions in order to note the effects of systematically controlled change.' However, the institutionalization of much evaluation has led to the development of a greater range of evaluation approaches in response both to the identification of a range of evaluation needs at different levels in organizations, and to pressures from inside and outside organizations for more substantial information to assist decision-making at all levels.

30 / Programme Evaluation and Quality

In educational research, the implications of the context in which learning takes place for the way we look at the way students learn have long been recognized. With organizations, the model still holds. No provider operates in a vacuum. Just as we need to look at learners' personalities and the wider environment in which they live and work in order to understand their study behaviour, so we need to be aware of the ethos of individual organizations, and the external environment within which *they* **must operate in order to appreciate their organizational behaviour. In the next chapter, we shall be looking at the wider environment in which providers of open and distance learning have to operate and at some of the implications of current trends for programme evaluation.**