It is very clear that museums are competing for customers in a relatively static market. The average number of visitors per museum reporting has steadily declined during the last decade from 72,000 in 1978, falling to 61,000 in 1982, 51,000 in 1986 and 48,000 in 1988. The average number of visitors per museum remained at 48,000 in 1990. During the period from 1978, museums reporting data to the English Tourist Board rose from 716 to 1,222. Museums surveyed by *Sightseeing in the UK* in 1990 numbered 1,547 (British Tourist Authority, 1991). In part, the fall in the average number of visitors per museum is accounted for by the growth in number of museums. In part too, the average figure is distorted by the accommodation of huge visitor losses at those national museums that have introduced charges – for example, admissions at the Natural History Museum were 47 per cent lower in 1991 than in 1981, and at the Science Museum were 53 per cent lower (Eckstein and Feist, 1992: 73). However, it would appear that it is not going to be easy to sustain visitor numbers, particularly in a period of recession.

Visitor research is an essential management information tool. It should include both qualitative and quantitive research, and be carried out as part of a systematic and planned programme. In the past, museums and galleries have seen this as expensive and time-consuming, but as museums strive to get closer to their audiences, the need to know first who they are, and second, what they think, will become more and more imperative.

4

Research and evaluation

in: Eileen Hooper-Greenhill, *Museums and their Visitors* (Routledge, 1994) 69-83

Research and evaluation are often confused. When is a piece of research an 'evaluation' and when it is something else? Evaluation as a concept has been used in museums as a blanket expression to refer to investigation and analysis carried out before, during and after a process, which has until recently often tended to be the exhibition process. Thus 'evaluation' might refer to work that is exploratory in nature and open-ended in its focus, such as for example general attitudes to themes for exhibitions, but might equally refer to work carried out to test detailed specific ideas, such as words and images for text panels.

Korn has suggested that 'evaluation' is the systematic collection of data and information about the characteristics, activities and outcomes of an exhibition or public programme (educational or leisure session, event) that is useful in making decisions about the programme's continuation or improvement. 'Research', on the other hand, involves the generation of new knowledge and the exploration of hypotheses, which, while not necessarily providing immediately useful information, does offer material for the development of theories (Korn, 1989: 221). Both processes use the same methodologies: questionnaires, interviews, focus groups and observations, and, although Korn's definitions are useful, in practice it is sometimes difficult to separate the two concepts. The major difference is in the objective of the work, in that evaluation is driven by the need for information for specific action in the short term, while research is stimulated by the need to know more for professional or personal awareness and for the development of conceptual frameworks (Munley, 1992).

Munley (1986) has identified five purposes for what she calls 'museum audience studies'. These are: *justification* of the value of the institution itself, or of its exhibitions or public programmes; *information-gathering* to aid in long-term planning; *assistance* in the *formulation* of new exhibitions or programmes; *assessment* of the effectiveness of existing exhibitions and programmes; and increased general understanding of how people use museums through the process of research and the *construction of theories*. The first two reasons, justification and information-gathering, require marketing and demographic studies; the next two, formulation and assessment of programmes, require

evaluation work; and the last reason, that of developing a general understanding of the use of museums, requires on-going research (Korn, 1989).

Munley's five categories represent a logical and coherent framework for the understanding and development of audience research and evaluation within museums and galleries. It is extremely rare that any museum or gallery has such a well-thought-out approach to audience research, and very unusual therefore that any one piece of research or evaluation can be placed within such a context. In many museums, very little research is carried out at all; in some, some of the processes identified by Munley might be in place from time to time, as circumstances demand. It is helpful therefore to see the nature and extent of a structured framework.

In the last chapter we looked at a range of pieces of research that related to the development of a general picture of the use of museums. Some of these had been carried out by museums and some by other agencies. Most of the work was discussed in the context of three of Munley's categories: the justification of the value of the institution, information-gathering for long-term planning and an increased general understanding of museum use. This chapter will examine the work that may be undertaken to justify the value of specific exhibitions or public programmes, and methods used in the formulation and assessment of exhibitions and programmes. We will begin by tracing the development of exhibit evaluation in Britain, review some of the types of evaluation that can be undertaken and conclude by considering the origins and uses of the two major theoretical approaches to evaluation.

In common with many museums across the world, neither research nor evaluation has been embraced with great enthusiasm by museums and galleries in Britain. Many museum workers have failed to see the relevance of researching or evaluating their products, and have viewed these ideas with distrust, as something which at best is time-wasting, and at worst, demonstrates a lack of professionalism, skills and abilities. Where more far-sighted curators, educators or designers have wished to carry out research or evaluation, it has sometimes been difficult to persuade the museum's governing body of the need to produce the necessary funds.

Evaluation studies are relatively well-established in the United States, where project funding often requires an assessment of work done in order to demonstrate value for money and the success of the project to the funder. Evaluation is also familiar in museums and science centres that depend on interactive exhibits, where it is part of the development process of the exhibit. Where science centres have a strong position in the museum field, as in India or some parts of South America, the relevance of evaluation studies for other types of museums is more in evidence.

Where work has been done in Britain, the result of research has suggested important modifications and improvements to original intentions and plans (MacDonald, n.d.). In one or two instances, research and evaluation have been built into a project from the beginning, and the development of the project has

been fully informed by continuing research. Examples are Gallery 33 at Birmingham Museum and Art Gallery (Peirson-Jones, 1992), the Natural History Centre (Greenwood, Phillips and Wallace, 1989), and the Interactive Technology Centre (Harlen, Van der Waal and Russell, 1986) both at the National Museums and Galleries on Merseyside. The success of those projects on Merseyside that have been based on thorough and on-going empirical research, and consistent trialling and evaluation of exhibits, has led to this research-based exhibit development process being used in many other areas of the work of the museum (Rees, 1990).

As some of the more innovative and exciting museum projects of recent years are seen to have been based on research and evaluation, so professional interest in the methods used is growing. Methods are also changing, to reflect improved understanding of the nature of museum and gallery communication and visitor motivation.

Early approaches to evaluation in Britain

The development of the theory and practice of evaluation in Britain acts as an interesting case-study. As it has been mainly limited to the work of one group of people working in one museum and has been thoroughly documented it is possible to trace the evolution of the ideas that lay behind the practical methods that were employed. Both the ideas and the methods have changed as time has passed and work has demonstrated the strengths and weaknesses of different approaches. An examination of this process of trial and error is useful and instructive for those who are considering embarking on a similar journey of discovery. As all museums begin to move closer to their visitors and to develop new ways of discovering and responding to their needs, lessons from the pioneers are of value.

The Natural History Museum in London was the early pioneer in the field of exhibit evaluation in Britain (Miles, 1986a, 1986b, 1988; Miles and Tout, 1991). In the early 1970s the museum drew up long-term plans for the development of the permanent galleries. The galleries were to be educational in approach, and the exhibitions were to be designed with educational principles firmly in mind. A further aim was to develop a systematic approach to the design of exhibitions, a museum technology.

The educational principles used were drawn from educational technology and programmed learning, in part based on the experience of designing distance-learning materials for the Open University, but also informed by work that had been carried out in the United States (Alt, 1977; Shettel, 1973; Screven, 1969, 1976). Educational technology works with a specific focus on the materials to be produced and pays much less attention to the needs of the learner. 'It is not possible to consider individuals even if we limit ourselves to the serious learner. Our predictions must concern populations of visitors' (Miles and Tout, 1979: 215). Evaluation of the effectiveness of the museum technology was built into the project from the beginning.

By the end of the 1970s a theory of exhibition design was presented (Miles and Tout, 1979). Based on the precepts of programmed learning, the following principles for effective exhibits were proposed: exhibitions should have explicitly stated objectives; the order in which material was to be learnt should be carefully considered and carried through to the exhibits; the material was to be arranged in steps of appropriate size; provision was to be made for various levels of ability, knowledge and interest; opportunities were to be offered for visitors to respond actively to the exhibit and to obtain some feedback. Exhibitions were to be designed so that the emphasis was on the concepts to be communicated rather than on objects *per se*, and exhibitions were to be presented in such a way that both the intellectual and the physical structures were made very plain to the exhibition visitor. It was assumed that exhibitions (at least in science museums) were a teaching medium and that visitors would be disposed to learn if the design of the exhibition was appropriate and skilful in enabling learning. As exhibitions were designed, this general theory was applied to specific cases (Miles, 1986b).

The main methods of evaluation used in the early days were interviews and observations (Miles, 1982: 159). Large-scale visitor surveys were used on an annual basis to research the nature of the audience and the opinions of visitors to the new displays and to the museum itself (Alt, 1980, 1983). The new displays were subject to much criticism at the time, but the research demonstrated that although most of the visitors to the museum did not have a scientific background, they found the new displays 'thought-provoking', 'made learning easy' and made 'the subject seem exciting' (Griggs, 1990).

Observation studies entailed watching visitors using the galleries, tracking the routes taken, looking at where stops were made, and measuring how long these stops lasted. These are difficult observations to make, and it is convenient if a gallery overhangs the exhibition to enable the tracking to take place unobserved. Observations can be made, however, of both 'cued' and 'non-cued' visitors, although some would argue that once a visitor is made aware that they are the subject of a tracking study ('cued'), their behaviour necessarily changes (Lucas, McManus and Thomas, 1986).

In the early observation studies at the Natural History Museum in London, two measures based on American research were used to assess the effectiveness of exhibits – 'attracting power' and 'holding power' (Griggs, 1984: 415, Miles; 1982: 159). 'Attracting power' is measured by expressing the number of people stopping at an exhibit in relation to the number visiting the exhibition as a whole. A display case where nearly everyone stopped would be successful on these terms. 'Holding power' is measured by noting the length of time that people remain at the exhibit. If the display case where nearly everyone stopped only attracted them for a few seconds this exhibit would not demonstrate much 'holding power'.

These methods used at the Natural History Museum were based in part on theories of exhibit evaluation that had been developed in America (Weiss and Boutourline, 1963; Parsons, 1965). This American approach was grounded in behaviourist psychology, with an emphasis on studying how people behaved in specific environments. The evaluation of exhibitions was built around the development of behavioural objectives, that is, the study and measurement of the behaviour of visitors in relation to specific quantifiable targets. Two main methods were used to measure visitor behaviour, observation and pen and paper tests. Interviews of visitors were also sometimes carried out. The first of these main methods, observation, was the study of what visitors did in exhibition galleries, with an (apparently) objective and distanced observer watching activities, movements and routes taken through the museum spaces. Movements were coded onto plans of the galleries. The success or failure of exhibits was judged from these tracking studies and observations, mainly in terms of attracting and holding power.

The second measure of exhibition success used pre-visit and post-visit pen and paper tests to assess what the visitor had learnt. This method assumed that the exhibition had set out to 'teach' something specific, and entailed the development of specific behavioural goals for visitors. 'Goal-referenced evaluation', therefore, consisted of measuring the success of the exhibition against specific stated objectives.

Screven (1986) gives us an example of a typical exhibition goal statement: 'To improve the general public's sensitivity to the influence of Greek pottery on Roman pottery design.' This is the type of objective that any exhibition developer might come up with. However, in order to assess performance in relation to such an objective, the objective would need to be restated in such a way that performance, or behaviour, could be measured.

> Given six pairs of colour slides of pottery, presented one slide at a time, each pair showing one Greek and one Roman piece, the visitor will correctly identify the Greek (or Roman) example in five out of six pairs.

The objective restated in this way specifies what the visitor does (identifies pottery), the conditions under which it is done (looking at slides), and the acceptable performance level for success (five out of six).

There are specific problems with this approach to evaluating exhibitions. Visitors are, on the whole, observed rather than talked to. Observation might indicate where people stop, but cannot show why the stops occur. What is it about the exhibit (or is it something else?) that caused people to stop? Unless researchers ask visitors why they stopped, or listen very closely to what people are saying to each other while they stop, it is impossible to answer this question. Assumptions are made about the success of the exhibit based on very limited evidence.

The attention through pen and paper tests to knowledge gain is often focused in a very specific and limited way, and suggests that only the prescribed way of relating to the exhibition counts as appropriate, which is clearly unreasonable. Visitors are assumed to be passive and mechanistic, and able to be manipulated by the exhibition rather than purposeful in their use of exhibitions according to their own agendas. The focus on specific limited objectives ignores both the complexity of the learning process, and the specific characteristics of

visitors. In addition, the learning objectives are considered in relation to cognitive gains only and ignore affective or emotional aspects (Griggs, 1984; Prince, 1984: 426; Roberts, 1990).

The exhibition team at the Natural History Museum rejected the use of behavioural objectives, and developed the less prescriptive concept of 'teaching points'. However, their reliance on exhibitions based on theories of educational technology led to a greater attention to the structuring of the material to be learnt (the exhibition) rather than to the motivations or interests of the visitor. Attention to the principles of communication that informed the emerging theory of exhibition design led to a lack of attention, as the theory was tried out in practice and exhibitions were built, to the perceptions of visitors.

Once exhibitions began to emerge, the responses of visitors could be explored (Miles, 1986b). Gradually the attention to the exhibition structure was balanced by an attention to the needs and responses of visitors (Miles and Tout, 1991). Both are, of course, necessary. Research, at the Natural History Museum and elsewhere, has revealed the importance of the social aspects of museum visiting (McManus, 1987, 1988, 1989), and the power of museums to change perceptions and attitudes rather than instil facts (Pond, 1983, 1984, 1985). The work of McManus, who developed ways of listening very closely to visitors as they stopped at exhibits, has been seminal in demonstrating the agendas that visitors bring with them to museums, but also the willingness that exists to 'talk' to the exhibit developer through interaction with the exhibit text (McManus, 1990). At the Natural History Museum, the experience of the whole gallery is now considered, rather than the success of individual exhibits, and attention is paid to the affective as well as the cognitive aspects of the museum experience (Miles and Tout, 1991).

The original approach to evaluation at the Natural History Museum was to build the exhibition, observe how it worked and then to use the results to revise the poor exhibits and to inform future work (Miles and Tout, 1991). This presented a problem. Any change that was required once the exhibition was mounted and opened, such as redesigning a display case that was being ignored, or repositioning it to make it more accessible, was bound to be costly both in terms of funding and in relation to the emotional investments of the exhibition developers and designers. Evaluation needed to be moved from the end of the exhibition process to a much earlier stage and needed, in fact, to be completely integrated into the process (Miles, 1985).

Front-end, formative and summative evaluation

Evaluation can take place at three main stages of the exhibition process, but these do, of course, overlap (Griggs, 1992). Front-end analysis is carried out at the beginning of the development of the ideas for the exhibition; formative evaluation tests ideas and exhibits while in production; and summative evaluation examines what has been achieved at the end of the process. Front-end analysis, also

sometimes called preliminary research, could be as simple as the testing of the titles for an exhibition, or might be as complex as a research study of visitor attitudes to the theme of an exhibition. Formative analysis generally consists of testing mock-ups of display panels or texts, or, in the case of exhibits in a science centre, trying out the various designs that might be possible for one of the exhibits. Summative evaluation takes place after the exhibition has opened, at the end of the exhibition production process, but during the running of the exhibition. The findings at this stage might lead to small-scale changes, in the lighting perhaps, but will on the whole become part of the front-end analysis for the next project. This circularity of the evaluation process makes a continuous use and attention to evaluation necessary. There are very few museums employing all of these methods in a planned way. One museum that has used evaluation consistently over a period of years is the Royal Ontario Museum (Lockett, 1991).

Front-end analysis aims to identify and eliminate errors which might occur in exhibitions before the detailed exhibition planning begins. An example from the Natural History Museum demonstrates how the approach to an exhibition altered following the opinions of the target audience. The exhibition British Natural History was designed to appeal to amateur naturalists. This group was defined as those people who answered yes to at least two of five questions concerning their level of involvement in natural history. Visitors to the museum were asked if they belonged to a naturalist club, read two or more books during the year on the subject or owned appropriate equipment (Griggs, 1984: 418). Fourteen of those people who fell into this category were then interviewed for thirty to forty minutes, during which time they were asked about their preferences in relation to the proposed exhibition. One of the questions they were asked was whether they would prefer displays arranged in taxonomic groups, or in habitat groups. The exhibition developers had wanted to make a display based on taxonomic groups, but the visitor preference was for habitat groups. The process of specific hypothesis testing happened at an early enough stage in the process of developing the exhibition to allow major decisions to be taken in an informed manner before resources had been committed.

Front-end analysis can be used in relation to specific exhibitions but can also be used more generally. At the Royal Ontario Museum in Canada 140 visitors were asked how they felt about the way a range of objects (piece of furniture, piece of sculpture, small decorative art object) were displayed. The visitors were carefully selected and variables such as age, level of education, type of interest and so on were taken into account in the analysis. The interviewees were asked to rank four two-dimensional mock-ups representing four typical exhibit techniques in order of preference. The posters showed:

- the object alone with its artefact label;
- a range of similar objects, each with its own artefact label;
- the object in a context of a room setting or vignette with a brief general description;
- the object within a storyline or thematic presentation which included layered text, graphics and artefact labels.

The results showed a distinct split between the last two methods of display, which were preferred in all cases, and the first two, which were ranked third and fourth. This was the case even for those people with fine-art-related hobbies and for the piece of sculpture, which indicates a very consistent preference for informational contexts for objects, including those that are often displayed in aesthetic contests (Lockett, 1991).

Preliminary research can be used to indicate what contexts and which information about objects might be found useful. Rose Kerr, the curator of the T. T. Tsui Gallery of Chinese Art at the Victoria and Albert Museum in London was fortunate in having a sponsor for the gallery that was enthusiastic about the idea of research into an appropriate mode of display. The curator and her colleagues had already spent some time watching visitors and acting as information points in the old Chinese gallery and in the newer Chinese Export Art Gallery. Log-books were kept that documented the questions that visitors asked. One of the things that became clear was that many people had found the labels difficult (Rose Kerr, personal communication). 'Goal-free' evaluation of this kind can be used to discover problems that have perhaps not been identified beforehand.

Work began on the development of the Tsui gallery in 1988, three years before it opened. This allowed sufficient time for a number of consultative measures to be taken. Several seminars were held in the gallery with a range of both museum staff and colleagues from outside the museum including historians, teachers and other museum professionals. These seminars were chaired by a facilitator, and general thoughts were shared. It was decided that the gallery would not be organised chronologically, but by themes. But which themes would be appropriate?

A small in-house questionnaire was carried out by a member of the Museum's Far Eastern Department in the Spring of 1989 to test visitor reaction to initial ideas; this was followed by a visitor survey carried out by National Opinion Polls. The questions visitors asked about the objects were confirmed as: 'What is it made of?'; 'Where, when and by whom was it made?'; 'How did the object enter the collection?'; and 'What does the decoration mean?' These questions were adopted for the introductory cases for the gallery, and also for the chapters in the accompanying catalogue (Kerr, 1991).

Small informal focus group discussions with people representative of specific target groups were held. Focus groups are small groups of people with similar characteristics, who meet for an in-depth discussion with an experienced moderator who uses questions, cues, probes and stimulation material to elucidate attitudes, opinions, reactions, issues and expectations in relation to a specific topic (Rubenstein, 1988). This is a recognised market research technique.

The groups discussing the Tsui gallery included 15- and 16-year-old boys, and girls of the same age, both of whom were studying for the General Certificate of Secondary Education; their teachers; teachers and parents of 8–11-year-olds;

teachers and members of the Chinese community in London. These groups were chosen because they were not frequent current users of the museum, but were groups that the museum wished to encourage. The idea of themes was welcomed, and most people agreed that more information should be provided than had been available in the old gallery. Most people were interested in how the Chinese objects were used and by whom, and this has been adopted as a major theme of the gallery. The idea of objects to touch was greeted as a good idea, and the curator consulted Shape, an organisation working with the arts and people with disabilities. Queries about the use of videos and interactive media in the gallery evoked a mixed response and worries about intrusive noise: in the gallery small-screen silent videos with a number of short sequences have been used.

The old Chinese gallery was organised chronologically, with artefacts spanning the period 3000 BC to the present day. It was presented according to the concerns of the curators and with little attention to the needs of the visitors. Research into visitor perceptions and interests has resulted in the new gallery being organised around six main topics: Burial, Worship, Eating and Drinking, Daily Life, Ruling and Collecting. The use of objects is indicated where possible, partly by discreet contextual arrangements.

There are two wonderful objects to touch, plans explaining the layout of the gallery and seating arranged as an integral part of the display to allow a comfortable pause. Some subject specialists have found the gallery altogether too different to be acceptable. Summative evaluation carried out shortly after the gallery opened tested the reactions of the visitors. The findings of two studies were very encouraging in relation to the general approach of the display, and useful practical suggestions were made about, for example, the siting of gallery plans.

Formative evaluation involves testing sections of the exhibition while in the process of developing the exhibition (Griggs, 1981; Jarrett, 1986; Screven, 1988). In practice this is very difficult, and is time-consuming, although there are some museums that use this approach on a regular basis (McNamara, 1988). In addition, it presupposes enough lead time to try out a number of the design elements of the exhibition, which in many situations is not possible. There are two areas where formative evaluation can be particularly helpful, and these are in the writing of texts and in the development of interactive exhibits.

Texts can be tested in a number of different ways. The vocabulary can be reviewed for unfamiliar and polysyllabic words, and 'to see if the intended message is being conveyed. Initial checking can assess the structure, complexity and length of the sentences (Coxall, 1992). If texts are read aloud, this often gives a clue as to how the language flows, which is important for communication. Family, friends, teachers, children, security staff and other colleagues can be asked to read the words and to describe what they understand. Many of these methods were adopted in a fairly informal way by the curator of Gallery 33 in the Birmingham Museum and Art Gallery. At the Children's Museum, Boston, trial labels are tested in a special display case,

where it is made clear that an evaluation is being undertaken. The labels are hand-written with a large pen on specially prepared sheets, which are blank except for a coloured edge. The curators are able to observe the public interacting with the display and to listen to the way in which they read the labels.

Other aspects of text that need consideration are placing and size. Does positioning affect the meaning of the words and how they are understood? A rough mock-up of part of the exhibit has been used at the Natural History Museum. Visitors were asked to spend a few minutes looking at the mock-up, and were then asked what they understood (Griggs, 1984: 416–17).

Interactive exhibits demand formative evaluation. Those things that are designed for people to use, to handle, to sit on or to manipulate must be tested. If this is not done, breakages and breakdowns are inevitable. Some designers of interactive exhibits would see the whole life of the exhibit as being one long process of formative evaluation. An example from the Franklin Institute in America describes how an exhibit involving ropes and a pulley failed because instead of one child pulling on each of ten ropes to test which made lifting a weight of 500 pounds easy, twenty-five children were all pulling on the ropes at the same time. The number of the ropes was reduced to four, guard rails were erected to narrow the entrance to the exhibit and a sign was erected that read: 'Can you lift 500 pounds by yourself?' (Borun, 1989).

Although this approach to designing exhibits is most often to be found in science museums, some art museums are also working with mock-ups of exhibits and texts. One example is the project at the Musée Picasso at Antibes, France, where formative evaluation is being developed with the help of an American consultant. The long-term aim is to develop enough expertise in audience advocacy in the staff of the museum as a whole so that the museum can act as a training ground for curators and other museum workers. This has grown from the concern of the French Ministry of Culture and Communication to increase the general educational effectiveness of French art museums (Screven and Giraudy, 1991).

Summative evaluation tests the exhibition after it has opened. In most museums where evaluation has been undertaken it is understood that summative evaluation, which judges exhibits after they have been produced, is perhaps less generally useful than formative evaluation that aids in the planning and development of exhibits. However, reviews of the effectiveness of exhibitions on completion can be helpful in the right circumstances.

An example of summative evaluation from the Discovery Room at the Royal Ontario Museum shows one way in which this approach can be used. The Discovery Room consists of modular units designed with handling, exploration and self-guided learning in mind. It has been the subject of fairly consistent assessment in terms of the actual use of specific exhibits and also in relation to the development of theories of discovery learning in the museum (Freeman, 1989: 41; Hooper-Greenhill, 1991a: 183). In 1984 a tracking study revealed that only approximately 7 per cent of visitors came as individuals, and all the others came in informal groups of between two and eight people.
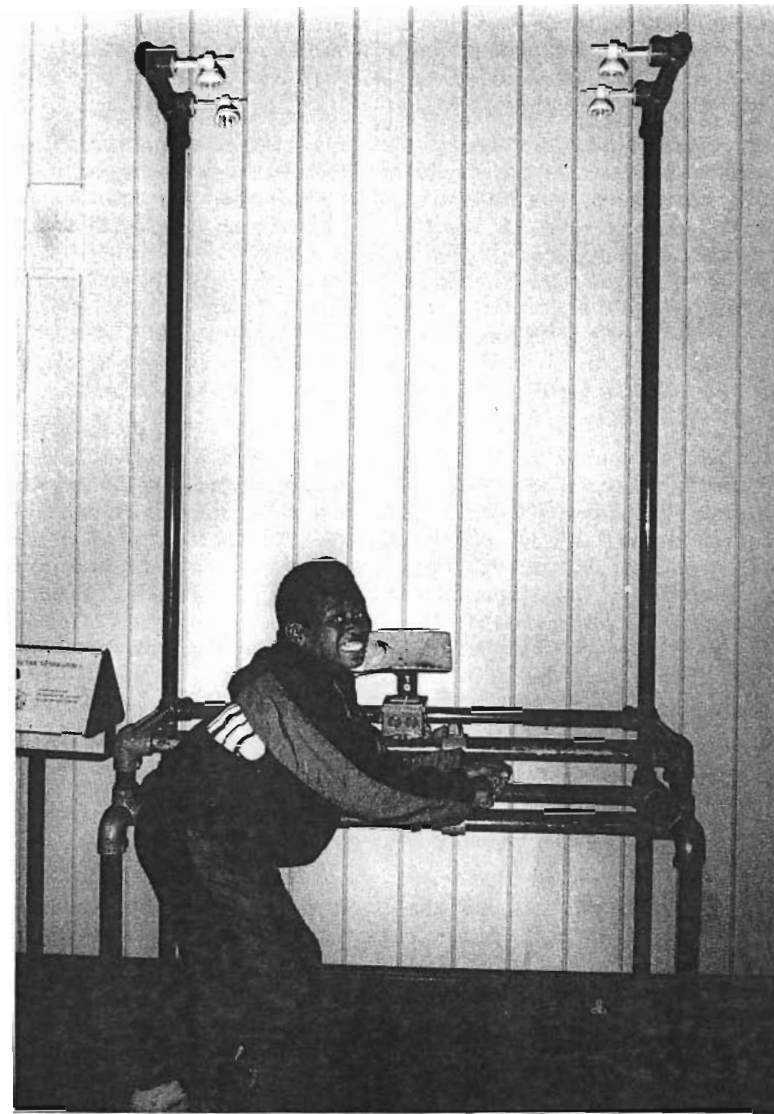

Plate 8 Interactive exhibits, such as this one where electricity is generated at the Museum of Science and Industry, Birmingham, demand formative evaluation. It is vital to test how the different parts of the exhibit relate to each other, to watch how people use the exhibit and to assess what level of wear and tear is likely. Failure to do this will result in an exhibition full of exhibits under repair!

Photo: Birmingham Museums and Art Gallery (Schools Liaison).

This observation led to the rearrangement of seating and spacing in the gallery to allow for more than one person per display. The tracking study also identified displays that were rarely if ever used. These displays were subsequently analysed and modified (Lockett, 1991: 41).

The Discovery Room at the National Museums of Scotland was the subject of a summative evaluation in the summer of 1990 (Stevenson and Bryden, 1991). The Discovery Room, which had been piloted in 1987 in the education room of the museum with a budget of £600, had become an enormously popular annual museum event with a capital outlay of £25,000 and an annual budget of £10,000 three years later. The Discovery Room is a space for a 'hands-on' experience, designed to encourage visitors to learn through and about objects and to appeal to people not accustomed to museum visiting. It is based partly on the work carried out at the Royal Ontario Museum and on Merseyside, and partly also on educational theory, but theory of a quite different nature from that used at the Natural History Museum in the 1970s.

The Discovery Room at the National Museums of Scotland derives from the tradition of learning by doing and learning through experience (Bruner,1962, 1966; Cole, 1985; Dewey, 1963). The concept of discovery learning can be traced back to the nineteenth century and beyond (Hooper-Greenhill, 1991c), where finding out about the world through the exploration of real things in real places is seen as one of the best ways of facilitating growth and development. Child-centred learning enables the development of skills and abilities within a structured, but open situation and has been a major element in educational thinking during this century. Currently, school pupils in Britain are being encouraged to find out for themselves through investigations of primary sources, through explorations of sites and environments, using active methods rather than passive book-learning. This approach to education is also to be found at the heart of theories of life-long learning, where people, places, activities and practical experience are all seen as opportunities for self-development throughout the lifetime of an individual.

The Discovery Room at the National Museums of Scotland broke new ground for the museum and it was important for the developers to demonstrate that their expectations concerning the success of the enterprise were in fact correct. A form of summative evaluation was designed using a method adapted from the Open University in Scotland, called Structured Video Recall. A video-tape of the eight themed areas of the Discovery Room was shown to visitors after their visit, and the ensuing conversation recorded. Although several methodological hiccups emerged during the process, the system worked very well and demonstrated that visitors enjoyed the experience so much that they wished for more discovery areas in the rest of the museum. Visitors had become interested in themes that they had not previously considered interesting, and had found answers to questions that had puzzled them for years such as 'How are teapot spouts made?' Some negative factors were identified: 'The doorway was very crowded and I could not see what was in the room.' The meaning that the experience had for individuals began to emerge. In relation to the Masks

area: 'Made me smile and use different expressions and voices. The image looked different from "me". I felt naughty!'; 'Not tempted to linger because I felt too self-conscious'.

The Discovery Room has been so successful in the main building of the National Museums of Scotland that it has found sponsorship funding to travel to different parts of Scotland and England. The Discovery Room exhibitions have been dismantled and re-erected in other museums, thus operating as an outreach and marketing arm of the National Museums, and enabling the development of new audiences for the museum outside the capital city of Edinburgh. Evaluation, in demonstrating in a researched and concrete way how the initial museum-based project worked, has been important in securing funding for the outreach developments.

Theoretical approaches to evaluation

Evaluation has begun to be of interest to museum workers, and is of particular concern to museum educators, although very little structured and documented work has been carried out to date. However, the emphasis on accountability and on museum reorganisation has spurred a need to find ways of evaluating the learning outcome of educational activities and events, in addition to the experience of exhibitions. Some curators are also now concerned to investigate the perceptions people have of museums, the attitudes they bring with them when they visit, and the meanings and values visitors attribute to what they see and do. The traditional methods of exhibit evaluation that developed in America do not offer ways of discovering answers to these questions.

Many evaluation studies from North America have been predicated on the idea that exhibitions are an instructional medium, that is, that they are didactic or educational. The view of education on which this attitude is based is very narrow. Education is equated with instruction, rather than with growth, and success is defined as measured against predetermined exhibit-related goals or objectives, as we have seen, rather than in relation to the needs and interests of the learner, which are largely ignored. The evaluation methods which have been developed have focused on analysis, generally of the completed exhibition, through quantitative number-based judgements made against measurable objectives defined by the exhibition developer (or the evaluator!), with the aim of judging the success or failure of the exhibition.

This kind of analysis has its roots in the study of the natural sciences, and in behavioural psychology. The predominant analytical model is laboratory-based testing, with carefully designed experiments carried out in controlled situations in such a way that they can be repeated and verified. In applying these methods in the museum as the test environment, visitors are observed, measured, counted and compared, and the results coded into standardised categories. Responses are converted into numbers so that they can be analysed statistically. Quantitative data is generated from repeatable experiments that test hypotheses.

The epistemological model used by these evaluators is 'scientific', stemming from the physical, laboratory-based sciences, and primarily designed to measure observable phenomena. This way of understanding the world studies only that which can be observed, measured and quantified. Within this positivist epistemological model, feelings, attitudes and other intangible elements which are more difficult to measure, cannot be included.

However, other ways of studying the world offer other methods of charting it, and increasingly, these methods are being adopted to try to find ways of understanding how it is that people make sense of museums. These other ways of understanding the world relate, not to the 'hard', physical, laboratory-based sciences, but to the 'soft' disciplines of sociology, ethnography and anthropology. These disciplines are generally field-based, that is, they study people as individuals in their own surroundings rather than as anonymous subjects in laboratories. The methods that have been developed have been designed specifically to study people, rather than, as with behaviourist psychology for example, having their roots in methods designed to study animals or things, such as natural elements.

The field-based methods stress observation rather than analysis, and work towards qualitative rather than quantitive data. Documentation, in-depth interviews and descriptions of case-studies result in narrative accounts, where the responses of interviewees are often quoted, producing 'thick' descriptions of practice. These descriptions look quite different from the reports resulting from the 'scientific' methods, which generally contain figures based on the manipulation of statistics.

'Responsive' or 'naturalistic' evaluation is concerned to describe rather than to analyse, and to understand, rather than to explain. It emphasises the meanings that experiences have for participants, and demonstrates this by using the words of those who were part of an event. Rather than focusing on a narrow set of predetermined objectives, naturalistic evaluation is open to the emergence of spontaneous activities and expressions, and seeks to account for a range of different values and perspectives. Evaluators aim to work jointly as a team with those being evaluated, and to work from the inside towards a mutual understanding, rather than to make judgements from an objective external stance (Korn, 1989; Hein, 1982).

The naturalistic approach to evaluation is beginning to become known in Britain. In the United States it has been familiar for some time and a body of good practice has been built up (Otto, 1979; Wolf, 1980). Interestingly, naturalistic evaluation has been used to study and develop educational programmes as well as to study exhibitions. A long-term evaluation project at the Children's Museum of Indianapolis, for example, enabled the education team to develop its effectiveness by thinking more clearly about its goals as a department, and by considering in depth solutions to issues such as the balance of quality and quantity, the relationship of the museum to schools and the role of the educators in exhibition design. Specific practical information was also produced that related to the design of exhibitions, the structure of the schools

programmes and the profile of the audience for one aspect of educational provision (Otto, 1979).

The two paradigms, 'scientific' and 'naturalistic', although stemming originally from different worldviews and different disciplines (Korn, 1989; Fay, 1975), should not be seen as mutually exclusive in practice. There are times when each is useful. For example, we can use an open-ended naturalistic evaluation to explore initial perceptions in relation to a proposed new gallery, as the curator of the T. T. Tsui Gallery of Chinese Art at the Victoria and Albert Museum did, where seminars were held with colleagues, and discussion with members of potential target audiences took place in focus groups. This will yield generalised descriptive data, where patterns, trends and individual comments will be valuable. If, on the other hand, we want to know how many secondary school children used the education department in any one year, a simple counting exercise will tell us.

There are a number of methods text books that can be used to help develop an understanding of the possibilities, although none of them are written explicitly for museums and galleries. *Social Survey Methods: a Fieldguide for Development Workers* (Nichols, 1991), is an excellent easy-to-follow beginners' guide. Written for Oxfam workers operating on a shoe-string budget in under-developed areas, many of the constraints will be familiar to those working in museums and galleries. More formal text books are *Sample Survey: Principles and Methods* (Barnett, 1991), a technical guide to survey data collection and analysis; *A Handbook of Qualitative Methodologies for Mass Communication Research* (Jensen and Jankowski, 1991), which has a good introduction to qualitative and humanistic theory which is then applied to the analysis of the media; and *Research Methods in the Social Sciences* (Frankfort-Nachmias and Nachmias, 1992), which is a comprehensive account of the range of both qualitative and quantitative methods. Methods are also usefully outlined in Korn (1989).