

INTRODUCTION TO EVALUATION: THEORY AND METHODOLOGY

RANDI KORN

Consultant in Program Evaluation

INTRODUCTION

Evaluation and research in the museum environment are generally conceived of as systematic processes using various data collection methods, that are used to examine programs and exhibitions and their effect on the museum visitor. Engaging in evaluation and research in the museum setting shows a commitment to the audience, a quest for excellence, and a desire to understand and describe the museum experience. Evaluation and research provide information that can guide decision-making, clarify programming options, improve audience relations, and inform staff about how visitors think and behave in museums. Therefore, an essential component of administrative planning and the development of audience-driven programs is evaluation and research.

The objective of this chapter is to define the various theoretical and methodological approaches to evaluation and to provide a broad introduction and background to the field. The idea that a program should be evaluated is certainly not unique to museums. Many of the ideas and methodologies used by museum evaluators are borrowed from the fields of behavioral psychology, education, and the social sciences. In this article, they are described in their

original context and, when appropriate, are explained from the museum perspective.

BRIEF HISTORY OF EVALUATION

The history of evaluation in museums is strongly tied to the educational mission of museums. The first museums were associated with the social and educational elite. This class of people viewed the museum as a place of research and a repository for which objects were cared. While the public role of museums grew in the 19th century, the ties between the museum and the elite remained strong. If exhibitions and label copy were not understood by the public because of technical jargon or difficult concepts familiar only to specialists, it was the problem of the visitor, not the staff.

At the turn of the century, museums started to hire educators and recruit volunteer docents. Their job was to talk about the exhibitions and in a language that would be comfortable to the visiting public, and soon, paintings were organized into meaningful groupings that were more oriented toward instruction.

In the late 20's and early 30's, two psychologists from Yale, Edward Robinson and Melton (Robinson, 1928; Melton, 1933), pioneered some studies that examined the behavior of museum visitors in an effort to record the educational role of museums. They observed visitors and noted their movements through the gallery, where they stopped to view a painting, and where they exited the gallery. The end result was a map showing the most frequently visited areas. While this research did not necessarily draw conclusions about education in museums, it represented the beginning of visitor behavior research and evaluation in museums, and it identified several visitor-related topics that could be explored using observation. Essentially, their research showed that where and how objects and interpretive information were placed in a gallery determined whether it was viewed by the average visitor.

Melton and Robinson continued conducting research in museums into the 30's, but there was somewhat of a lull until the late 60's. At this time, museums were under pressure to transform their galleries from quiet halls to dynamic, educational and cultural institutions. More and more, museums were looking towards public money and government grants for support. Museum staff, in a sense, were held accountable for these grants and were concerned about the effectiveness of these funded programs. As a result, more attention was directed toward evaluation procedures. Today, evaluation and research in museums is expanded beyond Robinson's and Melton's concept. Theories and methods from the fields of

education, psychology, and the social sciences are being used by museum evaluators. Museum evaluators conduct studies in exhibit galleries and classrooms; they analyze audiences, and measure the effectiveness of a variety of programs including interpretation, publication, wayfinding and public programs.

EVALUATION OR RESEARCH?

The distinction between "evaluation" and "research" is an important one. Evaluation and research share the same methodologies — standardized questionnaire, interview, and observation — but the information is used in different ways. Evaluation is a systematic collection of information about the characteristics, activities, and outcomes of a program that is useful in making decisions about the program's continuation and improvement (Patton, 1986: 14). Evaluation can provide feedback to program designers and administrators about the successes and failures of that program. It viewed in this way, as part of a program development system, evaluation facilitates the refining of a program at each stage of its design and implementation (Weiss, 1972a: 16). Research, on the other hand, emphasizes the creation of new knowledge by testing hypotheses (Patton, 1986: 14). It may not necessarily provide immediate, usable information, but rather, may generate theories.

Munley (1986: 19) uses the term "audience study" to describe both evaluative and research efforts. She identifies five reasons for undertaking an audience study. The first two reasons — justification of an institution or program, and gathering information as an aid for long-term planning — require marketing and demographic studies. The next two — assistance in the formulation of new programs, and the assessment of the effectiveness of programs — require evaluation studies. Finally, the last reason — increasing the general understanding of how people use museums — requires ongoing research (Munley, 1986: 19).

In most cases, the theoretical and methodological ideas described in this chapter fill under the topic of evaluation rather than research. Thus, "evaluation" is the descriptor used throughout the text. This chapter is not a how-to guide for conducting studies in the museum. There are, however, many good sources available for reference. A bibliography, which will help the reader find such models, concludes the chapter.

TYPES OF EVALUATION: UNDERLYING PARADIGMS

According to Kuhn (1970), scientists organize and conduct their

research around a paradigm. He defines a paradigm as a group of related assumptions about the world which provides a philosophical and conceptual structure for studying that world. There is a long-standing debate about the advantages and disadvantages of the research methods associated with the two most widely-used paradigms in the evaluation field: scientific and naturalistic.

The scientific paradigm is objective and outcome- or product-oriented. The methods associated with it are quantitative and may include experiments, surveys, and systematic observation. The naturalistic paradigm, on the other hand, is subjective and process-oriented. The methods associated with it are qualitative and include descriptions based on observations and unstructured interviews.

Reichardt and Cook (1979: 16-17) point out that the characteristics of each paradigm are logically independent, the methods are not logically linked to either paradigm, and the characteristics are not logically linked to each other. They assert that the fundamental difference between the paradigms is the assumptions on which they are based. These assumptions are as follows: the scientific paradigm seeks results by rejecting or *confirming* theories using quantitative methods; and the naturalistic paradigm seeks results by *generating* theories using qualitative methods.

The fact that certain methods are associated with each paradigm is unfortunate because evaluators think they must choose one method type over another and are discouraged from integrating them. Although there is a strong association between paradigms and methods in practice, Patton (1975: 40) and Flitstead (1979: 42) believe that neither paradigm can satisfy all evaluation needs because each has different strengths and weaknesses. The results obtained by one method or paradigm are not necessarily inferior to another. Kuhn (1970: 17:18) supports this notion by saying that a paradigm "... need not, and in fact never does explain all the facts with which it can be confronted". The type of questions that the museum staff members have is perhaps the most important consideration. To explain this point, an appropriate example is provided by Patton (1975: 13): If you want to know how many paintings adults view in a gallery, or to measure the frequency of visitor interaction during a gallery tour, statistical procedures are quite appropriate. But, if you want to understand how those paintings affect the adults, or what their interactions mean, qualitative methods including in-depth interviewing or observation are suitable.

When an evaluator is deciding which methods are most appropriate to use for a particular project, those staff members who will ultimately use the results of the study should be consulted. It is crucial that the evaluator understand the perspectives of staff

members, their questions and concerns about the program and how they hope to integrate the results into program planning. As the above example illustrates, evaluation questions determine which methods are most suitable. The first concern of the evaluator should be to provide staff members with answers to their questions. If their questions happen to require the use of both quantitative and qualitative evaluation methods, the evaluator should not hesitate to integrate the two paradigms.

Question types also indicate whether the evaluation will be formative or summative. This distinction, introduced by Scriven (1977), identifies those evaluations that collect information for program development or improvement as formative and those that make a judgement about the effectiveness of a program as summative. Formative evaluation focuses on ways to improve a program during its development or at any point during its life (Patton, 1986: 66). Summative evaluation is used for making decisions about whether to continue or terminate a program. The questions associated with each evaluation type are as follows:

Formative

What are the strengths and weaknesses of the program?

How can it be improved?

What is working well and what is not?

Summative

Has the program been effective?

Should the program continue?

Did the program produce the desired outcome? (Patton, 1986: 65-66).

The first research study in an art museum, conducted in 1928, was a quantitative study. Although the study (Robinson, 1928) provided useful information about displaying paintings, the idea of systematically collecting information about visitor behavior in a gallery did not survive. There was a surge of interest in evaluation activity in the late 1960's and 1970's in science and natural history museums. At that time, museum exhibitions were introduced as educational mediums and their effectiveness in communicating concepts and information to museum visitors was often evaluated.

While some aspects of art museum programming and visitor experiences may be quantifiable, such as learning art historical facts and information, many art museum staff members have never felt comfortable with "learning" as the only goal of an art exhibition. They believe that the experience occurring between a work of art and a visitor is not quantifiable. Art museum educators have also been influenced by the naturalistic studies conducted in the art education field. They have, therefore, moved towards a greater use

of naturalistic evaluation. Both paradigms, however, are applicable to situations and questions asked by art museum staff members. Therefore, a description of both the naturalistic and scientific paradigms follows. Included under each paradigm description are the methods associated with it. This is done to clarify the philosophy of the paradigms rather than to accentuate the association.

NATURALISTIC PARADIGM

Naturalistic evaluation, often called the alternative paradigm, emerged during the 1960's when educators, in general, claimed that the traditional paradigm (scientific) was not serving them well. Naturalistic evaluators objected to the use of programmed tests to evaluate student achievement without consideration of the teacher or the setting in which learning occurred. The naturalistic approach originated in the social sciences and is supported by the belief that human behavior is influenced by the environment in which it occurs (Wilson, 1977: 249), and that individuals bring differing values and perspectives to a setting (Wolf, 1980: 40).

Naturalistic evaluation considers the needs of those for whom the evaluation is being conducted: the museum staff members. Wolf and Tymitz (1979: 3) speak of naturalistic evaluation in the context of the museum when they say "... it responds to the subtle and distinct variations in people's expectations, attitudes, beliefs and perceptions. Issues are identified, discussed and clarified by sampling program participants, administrators, exhibit staff ...". Because staff members participate in the initial stages of the evaluation, the evaluation results will respond to their needs. Naturalistic evaluators believe that they are taking a humanistic approach to understanding social situations and interactions.

Naturalistic evaluators usually collect their data by using ethnographic techniques. Ethnography is the description and study of people and places (Sommer and Sommer, 1986: 35-36). It is an approach to research that was originally developed by anthropologists and sociologists. Ethnographic techniques, which may include unstructured observations and in-depth interviews, generally produce data that is in the form of field notes. Observers write everything that they hear and see and interviewers record responses verbatim. Descriptive information using natural language can represent people's beliefs, perceptions, feelings, and levels of knowledge. It can, in a sense, provide a "slice of life" (Wolf and Tymitz, 1977: 7).

Naturalistic evaluators conduct their studies in the natural setting or "in the field". The design of naturalistic studies is not planned prior

to entering the field; rather the plan evolves as the evaluator collects information from interviews and observations (Wolf and Tymitz, 1979: 3). Although naturalistic evaluators are aware that theories and ideas preexist, they prefer to see what the data generates naturally (Flisstead, 1979: 38). The interviewer, for example, will pose broad, open-ended questions to staff members and visitors in an effort to piece together the processes and circumstances of a program. As a result of conducting unstructured observation and interviews with participants, the purpose of an evaluation becomes apparent, and a pattern of interaction and meaning appears.

Educators have adapted the concept of ethnography and developed several different strategies and techniques for conducting naturalistic evaluations. Though they differ in some respects, they all share the same ideas: the involvement of the evaluator in the setting as both an observer and a participant; an attempt to understand occurrences from the participant's perspective; and consideration of the environment in which behaviors and events are taking place (Knapp, 1979: 119).

One particular naturalistic evaluation strategy, known as responsive evaluation, is discussed below. Though there are many useful approaches, this one is presented because it is used by art education researchers and is potentially beneficial to art museum educators. Two ethnographic techniques — observation and unstructured interviewing — are also discussed. These techniques are useful for most types of naturalistic evaluation.

Responsive Evaluation

Responsive evaluation, introduced by Robert Stake in 1975, is a form of naturalistic evaluation that has a specific frame of reference. The thrust of this approach is that the evaluator responds to the issues and concerns as stated by the teachers, students, audiences, or other important parties (Stake, 1975: 34). In Stake's own words, "An educational evaluation is responsive evaluation if it orients more directly to program activities than to program *intents*; responds to audience requirements for information; and if the different value-perspectives present are referred to in reporting the success and failure of the program" (1975: 14) (emphasis added). Stake believes that the perspectives of human beings and social situations are various, and that responsive evaluation allows for these multiple viewpoints.

Responsive evaluation takes form after observing the program in action and interviewing and informally talking with those who are important to the program. The program itself will suggest responses, and the issues will emerge naturally. The evaluator uses

the responses to form an illustrative narrative that will contain the intricacies, moods, and impressions of the respondents. Stake does not believe that responsive evaluation is always the answer — there are several ways to uncover the benefits or failures of a program. Tests and other quantitative data-collection methods should not be ruled out, but method choices should be made on the basis of what is important to those interested in the program (Stake, 1975: 16). Stake realizes that the value of education is diffuse and that a test may not indicate the effectiveness of instruction. He says that a test may not indicate the purpose of the program of art education, "... it is sometimes the purpose and training for the staff, or parent, to provide artistic experiences and training for the intrinsic value alone" (1975: 16). Responsive evaluation allows for these less concrete intents to emerge and emphasizes the issues important to each program.

Observation Techniques

One of the best methods for studying natural behavior is through observations. Behaviors are recorded as they occur, and they provide the evaluator with a direct experience (Guba, 1979: 274; Guba and Lincoln, 1982: 192). Recording of behaviors can be done through note-taking, tape recording, or video recording. Validity of observation as a data collection instrument, however, is a concern of many evaluators. A valid instrument is one that measures what it claims to measure. Guba (1982: 192) responds to this concern by affirming that because there is no time lag between the occurrence of the behavior and its recording, there is a guarantee of validity. Validity can also increase with repeated or independent observations (Sommers and Sommers, 1986: 38-9).

There are two types of observation: non-participant and participant. In non-participant, the observer plays one role, that of the observer. Wolf (1980: 41) comments on the value of non-participant observation in the museum. "Observing what a person actually does during a museum visit and recording his own explanations and perceptions about the experience provide powerful insight into the question of museum impact". He identifies five types of observation data.

1. Descriptions of the physical setting, including descriptions of people. This information is based on the fact (what is seen) rather than impression.
2. Descriptions of actions and behaviors.
3. Recordings of discussions as they are stated by participants.
4. Indicators, known as traces, of past behaviors. These can include written materials such as policy statements, demographic data and newspapers.

5. Physical evidence of "wear spots". A space or piece of equipment can have worn spots that might indicate heightened use (Wolf and Tyrnitz, 1979: 6-9).

Participant observation is a style of fieldwork in which the observer is playing two roles simultaneously: that of the observer and that of a participant in the program under study (Guba and Lincoln, 1982: 190). Data, in the form of field notes, are unobtrusively collected by the observer who has, in a sense, become a part of the evaluation environment. The observer interacts and converses with the subjects; becomes familiar with their perspective, concerns and beliefs; and acquires an understanding of the experience and complexities of the situation (Dodge and Bogdan, 1974: 68). This process provides the observer with access to the group's emotional reactions. The observer becomes a data source and therefore is able to build on his/her own knowledge as well as that of other group members (Guba and Lincoln, 1982: 193).

Participant observation, however, has some limitations. For example, the observer has limited experiences and relationships with group members because he/she assumes only a single role within the group. Usually a large volume of data is generated, causing difficulties during analysis. Perhaps the most heavily-debated limitation revolves around the ability of the observer to differentiate between his/her own biases and prejudices and the actual attitudes and feelings of the group. Finally, the observer must think of the ethical concerns involved in participant observation because in some forms of participant observation, group members are not aware that they are being observed by one of the participants.

Participant observation is used as an evaluation method in the art education field. Sevigny (1978: 3-5) thinks that quantitative observation systems, to be described later, only represent behaviors and interactions "through the screen of the observer and they do not qualify through the screens of the participants". Sevigny believes there to be four levels of involvement available to the observer: complete participant, participant-as-observer, observer-as-participant, and complete observer. These levels describe the extent to which the researcher is involved with group participants. He uses these four vantage points in his own research and believes that data collected from all perspectives provide a richer understanding of what actually takes place in a situation.

Unstructured Interviewing

The purpose of interviewing in naturalistic evaluation is to "understand how people in a setting view the setting, to learn their

terminology and judgements, and to capture the complexities of their individual perceptions and experiences" (Patton, 1982: 166). The unstructured interview is most beneficial when those being interviewed have specific knowledge in which the evaluator is interested, when the evaluator wishes to pursue a subject in depth, and when he/she intends to discover rather than verify ideas (Guba and Lincoln, 1982: 166). The format is essentially unstructured — that is, open-ended questions are asked and the subjects answer in their own words. The evaluation problem and situation are defined by the subjects as they each react to the broad issues raised by the interviewer. The respondent, rather than the evaluator, leads the interview. (Guba and Lincoln, 1982: 156; Guba, 1979: 275).

During an unstructured interview, the interviewer is careful not to suggest responses. An open-ended question is used to raise an issue without implying right or wrong answers. Open-ended questions include "what if" questions, those that ask the respondent to react to a proposed ideal situation, devil's-advocate questions, and questions that try to uncover the original source of a problem or situation (Guba and Lincoln, 1982: 178). The tone of the questions and the reactions of the interviewer must remain neutral, although part of the interviewer's job is to be sensitive, listen to the respondent and, perhaps, to follow his/her responses with a request for more information. The process of using the respondent's answers to delve more deeply into a topic is called in-depth interviewing (Sommer and Sommer, 1986: 89).

The advantages of unstructured interviews are threefold:

1. A good interviewer can create a comfortable atmosphere for an honest discussion.
2. There is flexibility when asking questions in that a question can be rephrased if it is misunderstood by the respondent.
3. They provide rich, descriptive information (Warwick and Linger, 1975: 129).

There are, however, an equal number of disadvantages:

1. People's actions do not always match their words.
2. Unstructured interviews are time consuming to conduct and difficult and laborous to analyze.
3. The sample is often small.
4. And, the interview process is susceptible to bias because the interviewer may, without realizing it, encourage or discourage a particular response (Sommer and Sommer, 1986: 104).

Most often, naturalistic evaluators rely on two or more ethno-

graphic techniques to collect data, thereby enhancing validity. A technique called data triangulation is used to corroborate the data, and to establish validity (Filstead, 1979: 44). For example, the information collected from observations will be used in conjunction with information from interviews to confirm conclusions about the program being studied.

Focus Groups

The focus group is a form of qualitative research that first emerged in the 1930's when social scientists started to question the usefulness of traditional research methods (Krueger, 1988: 18). Specifically, they were referring to the standardized questionnaire which provides response choices that are predetermined by the researcher. In focus groups, which are generally composed of seven to ten similar individuals, respondents have the opportunity to comment on ideas that they believe are important, while the researcher has a less dominating role (Krueger, 1988: 19). The discussions that emerge in a focus group are generally planned and concentrate on a well-defined subject.

Focus groups are used in marketing research to provide manufacturers with insight about the thinking of consumers. Focus groups in museums are used to better understand the visitor and their perceptions about the museum, to plan new exhibitions and programs, and to test mock-ups of interactive exhibition components. A primary characteristic of focus groups is the use of group interaction to produce data and insights that would be impossible to find otherwise (Morgan, 1988: 12). A permissive environment is created by a trained moderator, or the researcher. The sessions can take place in a conference room or a room designed specifically for focus groups. These rooms are equipped with a one-way mirror and microphones so group interactions can be observed and heard by clients as they occur. Usually several different sessions are held, each with different participants, in an effort to identify patterns of opinion and perception. The data are usually transcripts that are analyzed to reveal insights.

The advantages to focus groups are that the moderator can probe to get meaningful data without a lot of time investment, and they provide an opportunity to explore topics and generate hypotheses (Morgan, 1988: 21). The disadvantages in relationship to other qualitative methods are that they are not conducted in a natural setting such as participant observation. In contrast to the individual interview, the researcher has less control over the interaction. When analyzing focus group data, the researcher must consider that responses and behavior occurring in a group are subject to

influence and may not be representative of individual behavior (Morgan, 1988:21).

Summary

Naturalistic evaluators believe that multiple realities exist in a social environment. The patterns of interaction that naturally occur in a setting and the perceptions that are provoked by an experience create these realities. The naturalistic evaluator strives to reflect these diverse perspectives by using several ethnographic techniques to collect data. Throughout the entire evaluative process and during analysis, the realities of the situation emerge. Through qualitative inquiry, beliefs, perceptions, and attitudes of people become apparent, and ideas, theories, and problems are identified.

SCIENTIFIC PARADIGM

The scientific paradigm, often referred to as the dominant paradigm, originated from the tradition of experimentation in agriculture (Patton, 1986: 182). Education evaluators borrowed the techniques from the basic natural science model which included quantitative measurement, experimental design, and statistical analysis. Parlett and Hamilton (1976: 142) explain the similarities between education and natural science evaluation: agricultural evaluation is the assessment of a change by examining whether or not a plant has met prescribed standards. Students (like plants) are pre-tested and then are exposed to different experiences or treatments. Change is then measured to indicate the efficiency of these experiences. These kinds of studies produce numerical data to which statistical techniques can be applied. This experimental methodology is often thought of as the example of "good science" (Patton, 1986: 182).

In the tradition of science, experiments test hypotheses; in education and program evaluation, they generally test prescribed objectives. Weiss (1972 b: 24) identifies the classic evaluation question as, "To what extent is the program succeeding in reaching its goals?" This question assumes that there are goals, that they can be identified, and that there is an activity that works to achieve these goals (Weiss, 1972 b:24-26). Quantitative approaches emphasize the measurement, comparison, and generalization of results between two program types (Herman et al., 1987: 19-20). The evaluation tools used to collect quantitative data are standardized; all individuals in the sample answer the same questions. The responses from which individuals choose are predetermined by the evaluator so behaviors, attitudes, and knowledge gained fit into standardized categories. The responses are converted into num-

bers so the data can be analyzed statistically. Quantitative methods make it easy to collect data from a large sample of individuals. A standardized test is usually used to collect data on knowledge and learning in educational evaluation.

The Experimental Evaluation Design

The basic evaluation design for quantitative evaluations is the experimental model on which there are several variations. The two most common, the "pretest-posttest control group design" and the "posttest control group design", are called "true" experimental designs because they provide adequate controls for all sources of internal validity (Tuckman, 1978: 130). Both designs use experimental and control groups composed of randomly-chosen people. The two groups should be identical in all respects except for the fact that the experimental group is exposed as part of the test, to a particular program (often called a "treatment") and the control group is not.

In the pretest-posttest control group design, both groups receive a pretest and posttest. Unfortunately, there is no control for testing effect; that is, it is difficult to see if a gain on the posttest results is due to the experience of the pretest. This disadvantage of the design may be offset by the fact that a pretest can indicate if the experimental and control groups start out with equal abilities, and can highlight any communication problems with the test (Popham, 1972: 62). This design can be quite effective if the evaluator is sure that the pretest will not sensitize the group. The test results from the two groups are compared to see if the treatment was successful. Figure 1 illustrates the pretest-posttest control group design. The program is successful if the experimental group (A and B) has improved more than the control group (C and D).

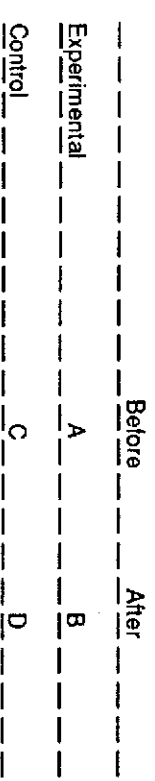


Figure 1. (Weiss, 1972 b:61) Pretest-posttest Control Group Design

The most useful "true" design is the posttest-only control group design. A pretest is not given, so the testing effect is controlled. This design is more cost-effective and can be used in most circumstances. For example, an evaluation can be designed to compare the effectiveness of two different teaching strategies with identical

objectives without altering the design. Figure 2 illustrates the posttest-only control group design with two experimental groups. If the difference between A and C is greater than the difference between B and C, teaching strategy 1 (experiment 1) is more successful than teaching strategy 2 (experiment 2). These kinds of comparative studies allow us to see which strategy or characteristic produces better results (Weiss, 1972 b: 78-79).

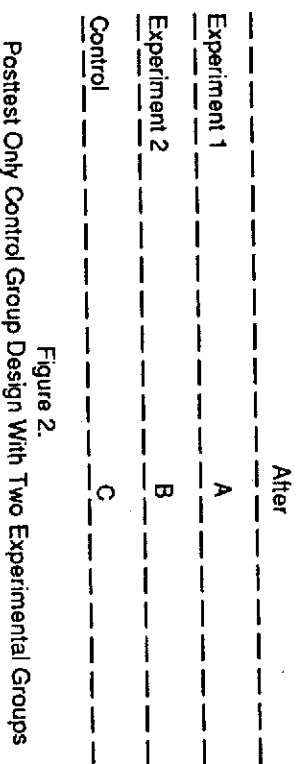


Figure 2.

Posttest Only Control Group Design With Two Experimental Groups

Experiments using independent and dependent variables suggest the cause and effect of a relationship. Variables are measurable characteristics such as age and sex, or the effect (as shown in test results) a treatment has on a group of individuals. An independent, or cause variable, is the one that is altered by the evaluator, such as a teaching strategy (see Figure 3). A dependent, or effect variable is that which has been observed and measured to see how it responds to the independent variable. In Figure 3, the dependent variable is knowledge gained as measured by 12th grader's test results. An independent variable is the stimulus variable and the dependent variable is the response variable. The purpose of an experiment is to determine the effect an independent variable has on a dependent variable. The analysis of program variables can explain why a program has the effect it does and specify which features are effective and which are not (Weiss, 1972 b:46). The evaluation question for Figure 3 might be, "Which teaching strategy produces more visually literate 12th-graders?"

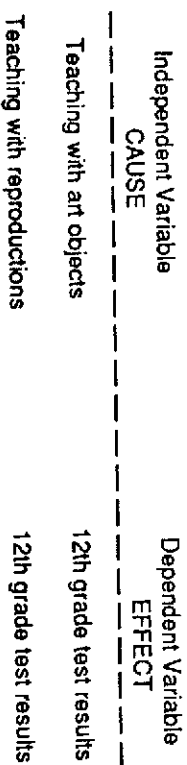


Figure 3. Independent and Dependent Variables

The experimental model is a data-collection procedure that is useful for finding out how well a program or particular teaching strategy has achieved its goals. If other types of information are needed, different kinds of evaluation techniques should be used. Surveys and systematic observation techniques can supplement the results of an experiment. Survey results alone provide a wealth of information about visitors' attitudes and expectations. The following sections discuss these two useful quantitative evaluation techniques.

The Sample Survey

The sample survey is one of the most widely-used data collection instruments in the social sciences. It is a systematic, standardized approach to collecting information from a randomly-selected sample of individuals (Rossi et al., 1983: 1). More and more the sample survey is being used not in isolation, but as an evaluation tool in conjunction with other methods (Warwick and Linger, 1975: 51). Although survey methods appear to be fixed, social scientists are continually developing new, more reliable procedures. The results from a survey can be used to describe and analyze audiences for marketing and membership purposes, to highlight the differences between visitors and non-visitors, and to provide additional information for an evaluation project.

An evaluator or researcher may choose to use a standardized questionnaire if a large sample is required, if the data will be collected over a long period of time, if his/her concerns about the audience can be written in the form of questions, and if the data will be analyzed statistically. Each respondent who completes a questionnaire will have answered the same questions in the same order. The responses to most of the questions are also standardized so that all the respondents use the same set of response options to answer. Although questions and responses are standardized, surveys can collect details pertaining to the attitudes, beliefs, behavior, and attributes of people. Each question is preconceived based on clearly stated goals that are project-specific. The general goal of any questionnaire should be to obtain information relevant to the purpose of the evaluation or research project and to collect information that is reliable and valid (Warwick and Linger, 1975: 127).

Surveys usually contain two categories of questions: open-ended and closed. Open-ended questions do not provide the respondent with any answers from which to choose. Rather, he/she is encouraged to answer in his/her own words. Although the responses are difficult and time-consuming to analyze for a large sample, open-ended questions are particularly useful in planning a

questionnaire. The answers help the evaluator see which responses would most likely provide the best choices for respondents. Open-ended questions are often used if the sample is small.

Closed questions are used most often on the standardized questionnaire. Each question includes a predetermined set of responses from which the respondent must choose the most appropriate answer. Within the category of closed questions, there are three types:

1. Questions with unordered choices — This is the most common question type. The respondent can choose from several predetermined alternative answers that are each independent of the next. The respondent is expected to choose one or more of the alternatives. For example:

If you were a visual arts major in college or graduate school, which one of the following best describes your main concentration?

- 1) Does not apply to me
- 2) Art History
- 3) Art Education
- 4) Aesthetics
- 5) Studio

2. Partially closed-ended questions — These are questions with unordered choices that also include an option for respondents to create their own answer by choosing the "other" option and filling in the blank. For example:

How did you find out about the Museum? (Choose the appropriate responses by circling the number or numbers.)

- 1) Word of mouth
- 2) Road sign on highway
- 3) Newspaper
- 4) Museum publication
- 5) Tourist Guidebook
- 6) Other (Please describe) _____

3. Questions with ordered choices — The respondent can choose from several options that are arranged on a scale. The scale can indicate values such as frequency, quantity, quality, and level of agreement and disagreement. These questions are also useful for exploring a specific, well defined concept. Respondents are encouraged to think about one idea in detail (Dillman, 1978: 89-91).

For example:

How often have you visited this Museum in the last year?

- 1) This is my first time

- 2) One other time
- 3) 2-3 times
- 4) 4-6 times
- 5) 7 or more times

Or:

For each statement below, circle the letter or letters that best reflects your opinion.

SA = Strongly Agree

A = Agree

U = Undecided

D = Disagree

SD = Strongly Disagree

SA A U D SD I like to view works of art without written or verbal information.

SA A U D SD I like to know why a work of art is important.

SA A U D SD I like to know what to look at in a work of art.

The answers to closed questions are easier to code and analyze and comparisons from one respondent to the next are simplified. There are, however, few opportunities for self-expression, and predetermined responses may suggest right and wrong answers.

There are several ways in which to administer a questionnaire: 1) mail survey; 2) self-administered survey and; 3) face-to-face survey interview. Mail surveys and self-administered questionnaires are advantageous for several reasons:

1. They are cost effective because there is little training involved in administering the survey;
2. And they are anonymous.

The disadvantages are:

1. The respondent has no one to turn to when he/she does not understand a question and may answer it incorrectly;
2. And, there is no one to encourage the respondent to complete the questionnaire.

If a questionnaire is to be self-administered, the construction of it as a whole, and of each individual question is crucial. The success of a self-administered questionnaire lies in the design of the questionnaire and in a well-planned distribution procedure. The Total Design Method (Dillman, 1978), if followed step-by-step has proven to achieve excellent results for the mail survey.

The advantages of the face-to-face interview are that the questionnaires are usually completed correctly and the interviewer can explain difficult questions or instructions. They are, however, costly to administer because of the expense involved in training interviewers (unless they are volunteers) or hiring professional interviewers and they are difficult to code and analyze. In addition, respondents

sometimes feel they must provide the interviewer with answers that are desirable or expected, rather than honest.

Systematic Observation

Posttest scores or visitor survey results are often influenced by behaviors which can best be tracked by conducting systematic observations in the evaluation environment. Systematic observations have a specific purpose and are well planned before the recording of behavior begins. A checklist of behavioral categories is usually developed after casually observing natural behaviors in the setting. The evaluator is then able to observe people and simply mark the occurring behaviors on the list, thus simultaneously, the behavior categories and the quantities are tested for reliability.

Observation results are compatible with survey and posttest scores. If visitors complete a questionnaire inquiring about their reaction to an exhibition, observing visitors as they tour the exhibition would verify and complement the survey data. Systematic observations, prior to conducting a survey, can provide background information for developing the questionnaire.

Summary

Qualitative or scientific evaluation emphasizes the relationship between the results of experiments and surveys with prescribed goals and objectives. The experimental design is an excellent evaluation model for examining one or more program strategies to see if goals have been met. The survey provides another method for collecting different kinds of information about programs and program participants. While the results of a survey provide measurements that can be compared to goals and objectives, they also provide background information and general knowledge about visitors that is useful for planning.

CONCLUSION

This discussion describes the primary characteristics associated with each paradigm. Qualitative evaluators understand there to be multiple realities that can be "discovered" during the evaluation process, while quantitative evaluators examine hypotheses against prescribed objectives. Qualitative evaluation emphasizes the patterns of interaction that occur naturally in a setting, and quantitative evaluators design experiments using groups of randomly chosen people.

Several data collection methods that may be useful in the museum setting are discussed within the parameters of each

paradigm. This, however, is not meant to support the dichotomy that exists between paradigms and method choice, but rather, to present a logical framework for learning about evaluation. Choice of methods should be dictated by evaluation questions and problems, not by preconceived ideas of the right or wrong paradigm. Decisions should transcend the boundaries that exist between paradigms. The objective of any evaluation is to provide information to staff members in a form that will answer their questions and assist them in planning. Staff members should feel a part of the evaluation process from beginning to end.

BIBLIOGRAPHY

- Dillman, D. A. (1972) *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley — Interscience.
- Dodge, M. and R. Bogdan (1974) "Participant Observation: A Promising Research Approach for Educational Technology". *Phi Delta Kappan* Sept. 1974: 67-69.
- Filstead, W. J. (1979) "Qualitative Methods: A Need Perspective in Evaluation Research." pp. 33-48. In T. Cook and C. S. Reichardt (eds.), *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills: Sage.
- Guba, E. G. (1979) "Naturalistic Inquiry." *Improving Human Performance Quarterly* 8 (4): 268-276.
- Guba, E. and Y. Lincoln (1982) *Effective Evaluation*. San Francisco: Jossey-Bass.
- Herman, J. L., L. Lyons Morris, and C. T. Fritz-Gibbon (1987) *Evaluator's Handbook*. Beverly Hills: Sage.
- Krapp, M. S. (1979) "Ethnographic Contributions to Evaluation Research." pp. 118-139. In T. Cook and C. S. Reichardt (eds.), *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills: Sage.
- Krueger, R. A. (1988) *Focus Groups: A Practical Guide for Applied Research*. Beverly Hills: Sage.
- Kuhn, T. S. (1970) *The Structure of the Scientific Revolution*. Chicago: University of Chicago Press.
- Melton, A. W. (1933) "Studies of Installation at the Pennsylvania Museum of Art" *Museum News* 10 (1): 8.
- Morgan, D. L. (1988) *Focus Groups as Qualitative Research*. Qualitative Research Methods (Vol. 16). Beverly Hills: Sage.
- Munley, M. E. (1986) "Asking the Right Questions." *Museum News* 64(3): 18-23.
- Parlett, M. & D. Hamilton (1976) "Evaluation as Illumination: A New Approach to the Study of Innovative Programs." pp. 140-157. In Gene V. Glass (ed.) *Evaluation Studies Review Annual* (Vol. 1). Beverly Hills: Sage.
- Patton, M. Q. (1986) *Utilization-focused Evaluation*. Beverly Hills: Sage.
- Patton, M. Q. (1975) *Alternative Evaluation Research Paradigm*. North Dakota Study Group in Evaluation. North Dakota: University of North Dakota.
- Patton, M. Q. (1982) *Practical Evaluation*. Beverly Hills: Sage.
- Popham, W. Jones (1972) *An Evaluation Guidebook*. Los Angeles: Instructional Objectives Exchange.
- Reichardt, C. S. and T. Cook (1979) "Beyond Qualitative Versus Quantitative Methods" in T. Cook and C. S. Reichardt (eds.) *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills: Sage.
- Robinson, E. S. (1928) *The Behaviour of the Museum Visitor*. *American Association of Museums Monograph* (New Series No. 5). Washington D. C.: American

- Association of Museums.
- Rossi, P. H., J. D. Wright and A. B. Anderson (1983) *Handbook of Survey Research*. Orlando: Academic Press.
- Scriben, M. (1977) "Methodology of Evaluation", pp. 334-371. In A. A. Bellack and H. M. Kliebard (eds) *Curriculum and Evaluation*. Berkeley: McCutchan.
- Sevigny, M. J. (1978) "Triangulated Inquiry: An Alternative Methodology for the Study of Classroom Life." *Review of Research in Visual Arts Education* (8): 1-16.
- Sommers, R. & B. Sommers (1986) *A Practical Guide to Behavioral Research: Tools & Techniques*. New York: Oxford.
- Stake, R. (1975) *Evaluating the Arts in Education: A Responsive Approach*. Columbus: Charles Merrill Co.
- Tuckman, B. W. (1978) *Conducting Educational Research*, 2nd edition. New York: Harcourt Brace Jovanovich.
- Warwick, D. P. and C. A. Lininger (1975) *The Sample Survey: Theory and Practice*. New York: McGraw Hill.
- Weiss, C. H. (Ed.) (1972) *Evaluating Action Programs: Reading in Social Action and Education*. Boston: Allyn and Bacon, Inc.
- Weiss, C. H. (1972) *Evaluation Research: Methods for Assessing Program Effectiveness*. N. J.: Prentice-Hall.
- Wilson, S. (1977) "The Use of Ethnographic Techniques in Educational Research." *Review of Educational Research* 97 (1): 245-265.
- Wolfe, R. L. and B. L. Tymitz (1977) "Toward More Natural Inquiry in Education." *CEDR Quarterly* (10): 7-9.
- Wolfe, R. L. & B. L. Tymitz (1979) *A Preliminary Guide for Conducting Naturalistic Evaluation in Studying Museum Environments*. Washington, D. C.: Smithsonian Institution, Office of Museum Programs.
- Wolfe, R. L. (1980) A Naturalistic View of Evaluation. *Museum News* 58 (6): 39-45.