

An Overview of The Methodology of Visitor Studies

Stephen Bitgood, Ph. D.
Jacksonville State University

Reliability and Validity

If we are to make important decisions from our research/evaluation results, our measurements must meet the standards of science. The two most important standards, "reliability" and "validity", will be discussed here.

Reliability

The term "reliability" refers to the consistency or stability of measurements. If two observers are observing the same visitor, we expect the same measurements to be reported by both observers (referred to as "interobserver reliability").

Reliable measurement depends upon two things. First, it requires objectivity (not allowing personal feelings or expectations to influence the measures). If behaviors to be measured are precisely defined, there should be less room for subjective measurement.

Reliable measurement also depends upon standardization. Each person using the measuring system should use it in the same way. Nonstandardized practices can make interpretation of results difficult. For example, in a survey, if a question is asked in different ways, the answers might vary depending upon the particular wording of the question.

Validity

Validity is a complicated concept that includes many subconcepts. A brief description of some of the subconcepts is below:

1. Assessment Validity

Assessment validity refers to the accuracy of conclusions about your measurements.

Content validity. Is your sample of visitor behavior representative of the population of behaviors you wish to test?

Predictive validity. Can you use a sample of behavior to predict the visitor's behavior to other exhibits?

Construct validity. Are your measurements really measuring the concepts you think they are? If visitors are able to answer multiple-choice questions correctly, does it mean that they learned this information from the exhibit?

Recording validity. Does your measurement system distort the actual behavior of visitors? For example, when visitors are asked to estimate time they spend at an exhibit or in a museum/zoo, they often overestimate time (e.g., Bitgood & Richardson, 1986).

Convergent validity. Degree to which a measurement device correlates with similar measures of the same concept.

Discriminant validity. Degree to which a device produces results different from other measurement devices when it should differ.

Ecological validity. Degree to which simulated environments, slides, or verbal representations of environments are related to the real environment through behavioral measures. Do these measurement devices relate to how people behave in the real environment?

2. Experimental Validity

Internal validity. Are the results due to the factors that you think? If visitors read one exhibit label more than another, can you conclude that it was the content that produced the difference rather than the number of words, size of letters, position of the label, etc.?

External validity. Do your results generalize or extend to other settings with other visitors?

Types of Visitor Research

There are several types of research used in visitor studies. These types differ primarily in terms of the control they exert over the variables being studied and, consequently, in terms of what kinds of questions they answer.

Experimental Research

General description. An experiment attempts to determine how isolated variables influence visitor behavior. For example, let us assume that we wish to determine how number of words on an exhibit label influences visitor reading. In an experiment we might present labels of varying numbers of words to different groups of individuals and measure whether or not they read and how long they read each label. If we conduct this experiment, we would probably find that the shorter the label, the more likely that visitors will read. If factors such as the characteristics of the visitors (age, gender, education, group size) and characteristics of label (content, letter size, distance from visitor) are the same from group to group, we expect that any differences in reading between one label length and another is due to the number of words rather than other, uncontrolled factors. Other variables (time of day, crowding, climate control, etc.) must also be considered and held constant.

An experiment attempts to establish cause-and-effect relations by showing that variables we manipulate influence behavior when other factors are held constant. Subjects are chosen carefully according to acceptable, scientific selection procedures. Those chosen must be representative of the total population of individuals about which we wish to draw conclusions. For example, a sample of weekday visitors may be quite different than weekend visitors since families are more likely to be represented on weekends while family members are likely to be in school or work during the weekdays.

Laboratory experiments. This type of experiment is conducted in a very carefully controlled environment. A laboratory setting allows much greater control over events than is possible in the real world. Thus, unexpected interruptions and intrusions can be carefully controlled and interpretation of results becomes more straightforward. However, the subjects in a laboratory experiment may realize that the situation is not realistic and may behave differently than in the real world. Thus, laboratory experiments may have less "experiential realism" than other types of research.

Field experiments. A field experiment is conducted in a real world situation. It is difficult to conduct such experiments since it is often impossible to exert enough control. For example, it is usually difficult to assign subjects to groups in an acceptable manner for experimentation.

Simulations. If the research cannot be conducted in an appropriate field setting for one reason or another, researchers may attempt to simulate the real world by creating important aspects of the setting. Museums have been simulated with slides and videos.

Quasi-experimental studies. When the assignment of subjects to groups cannot be controlled by the experimenter, it may be possible to use "quasi-experimental designs" (Cook & Campbell, 1979). These designs attempt to create controls as close as possible to regular experiments. Many visitor studies articles fit into this type (e.g., Bitgood, Pierce, Nichols, & Patterson, 1987).

Correlational Research

This type of research examines the relationships between visitor behavior and variations found in the setting. For example, Bitgood, Patterson, & Benefield (1988) measured visitors' behavior in zoo exhibits of similar species across 13 zoos throughout the U. S. They found that factors such as the size of the species, movement, and presence of infant were correlated with longer viewing times no matter where zoo visitors were observed. While correlational research does not allow the researcher to make strong conclusions about causal effects, the results may be suggestive of factors influencing behavior. For example, while no one may be able to

prove conclusively that cigarettes cause cancer, the correlation between smoking and cancer is strongly suggestive.

Descriptive/Observational Research

Descriptive or observational research gives us information about how visitors respond (either through direct observation or self-report) but this method does not allow us to make conclusions about how specific factors influence the behavior of visitors. This method merely describes how visitors behave, often in a qualitative rather than quantitative manner.

Methods of Measuring Visitor Behavior

Direct Observation

Recording what visitors actually do is a common way to measure visitor behavior. Usually this involves exhibit-related behaviors, but this is not always the case (see Falk, Koran, Dierking, & Dreblow, 1985). Falk et al. (1985) examined visitor attention to exhibit content, exhibit setting, and social group.

Recording can be obtrusive or unobtrusive. In unobtrusive recording, visitors do not know they are being observed. In many situations, people behave differently when they know they are being watched. Thus, studies using obtrusive recording may be more difficult to interpret. For example, Bechtel (1967) found that visitors spent longer in an exhibit area when they knew they were being observed.

Below is a list of some of the behaviors measured by direct observation.

- Visual attention to the exhibit and/or label (looking at or glancing at exhibit).
- Stopping and visually attending to the exhibit or label.
- Viewing time.
- Time in exhibit area.
- Pointing to some aspect of the exhibit.
- Touching or manipulating some aspect of the exhibit.
- Social interaction between or among visitors.
- Circulation path through an exhibit or facility.

These behaviors can be monitored throughout the museum or exhibit area (tracking procedure) or at a specific exhibit area (focused observation procedure). In the tracking procedure visitors are followed and their pathway and other aspects of their behavior are carefully recorded. In the focused recording procedure visitor behavior is recorded at isolated exhibits.

Behavior mapping. Behaviors are marked on a drawn-to-scale map. This method allows one to determine which specific behaviors occur and whether or not they are

associated with features of the setting (e.g., Ittelson, Rivlin, & Proshansky, 1970).

Other methods. Indirect measures (erosion techniques like worn pathways in the grass; leftover techniques such as pieces of litter) are occasionally used. In addition, photos and video/ audio recording are often used.

Self-Report Methods

Self-report methods include such techniques as questionnaires, interviews, focus group methods, and rating scales. Self-report methods by their very nature are "reactive" since the visitor knows he/she is being treated in a special way. Visitors may try to be "helpful" by exaggerating the pleasure of their experience or telling the interviewer what he/she thinks is expected. Any good textbook on research methodology in the social sciences will describe the pros and cons of self-report (e.g., see Marans, 1975). Also, see Loomis' (1987) chapter on the use of visitor surveys and Hood's (1986) paper.

Questionnaires. These are paper-and-pencil devices used to assess factual information and/or attitudes.

Interviews. Visitors are asked questions and their answers carefully recorded.

Focus groups. This technique uses a directed interview with small groups who are carefully chosen to represent some segment of a population of potential or actual users.

Rating scales. This method attempts to force respondents to rate the strength of cognitive or affective reactions to some aspect of the environment (e.g., physical feature, staff friendliness).

References

- Bechtel, R. (1967). Hodometer Research in Museums. *Museum News*, 45(7), 23-25.
- Bitgood, S., Pierce, M., Nichols, G., & Patterson, D. (1987). Formative Evaluation of a Case Exhibit. *Curator*, 30(1), 31-39.
- Bitgood, S., Patterson, D. & Benefield, A. (1988). Exhibit Design and Visitor Behavior: Empirical Relationships. *Environment and Behavior*, 20(4), 474-491.
- Bitgood, S. & Richardson, K. (1987). Wayfinding at the Birmingham Zoo. *Visitor Behavior*, 1(4), 9.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimentation*. Chicago: Rand McNally.
- Falk, J., Koran, J., Dierking, L., & Dreblow, L. (1985). Predicting Visitor Behavior. *Curator*, 28(4), 249-257.
- Hood, M. (1986). Getting Started in Audience Research. *Museum News*, 64(3), 24-31.
- Ittelson, W., Rivlin, L. & Proshansky, H. (1970). The Use of Behavioral Maps in Environmental Psychology. In H. Proshansky, W. Ittelson, & L. Rivlin (Eds.) *Environmental Psychology*. New York: Holt, Rinehart, & Winston.
- Loomis, R. J. (1987). *Museum Visitor Evaluation*. Nashville, TN: American Association for State and Local History.
- Marans, R. W. (1975). Survey Research. In W. Michelson (Ed.), *Behavioral Research Methods in Environmental Design*. Stroudsburg, Pa: Dowden, Hutchinson, & Ross.

Visitor Evaluation: What Is It?

Stephen Bitgood
Jacksonville State University

There are many concepts and issues related to evaluation discussed in the visitor studies literature. This article is a brief summary of some of these issues as I see them.

Research vs. Evaluation

Many writers (Friedmann, Zimring, & Zube, 1978; Patton, 1987; Screven, 1988) have made a distinction between "research" and "evaluation." Others (e.g., Loomis, 1988) see evaluation as a specific form of research. Below is a summary of some of the distinctions made by those who argue research and evaluation are distinct. [These distinctions are not universally accepted. I am among those who see little difference between research and evaluation.]

- Research attempts to control extraneous factors, while evaluation attempts to describe these factors.
- Research is concerned with discovering the causes for behavior; evaluation is concerned with factors that influence behavior.
- Research aims to reduce the number of factors; evaluation examines complex systems.
- Research uses rigorous methodology; evaluation is less formal.
- Research uses quantitative, statistical analysis; evaluation is more likely to be qualitative.
- Research requires highly trained professionals; evaluation can be conducted by those who have less training and knowledge.
- Research is expensive and time consuming; evaluation can be carried out quickly and inexpensively.

While these distinctions can be made in extreme cases, there are many studies (e.g., Loomis, Fusco, Edwards, & McDermott, 1988) that seem to serve both purposes.