Chapter 8:
Formative Evaluation:
Conceptions and Misconceptions

C. G. Screven
University of Wisconsin-Milwaukee
Milwaukee, Wisconsin

Introduction

Formative evaluation is a tool which gives you better information about what viewers are likely to do or not do, feel, etc. when confronted with prospective instructions, labels, objects, layouts, or topics. Such information contributes to decisions you must make in any case, with or without evaluation. How much time will visitors spend? What do they want (or need) to know about topic X? What misconceptions about topic X need correction? Are visitors able to follow instructions? What is an effective layout for achieving a desired impact? How should topic headings be worded? Where should directional signage be located?

There is much misunderstanding about evaluation, reflected, for example, by arguments over naturalistic versus goal-referenced evaluation, quantitative versus qualitative approaches, and objective versus subjective and person-to-person approaches to evaluating exhibit impact. Such arguments are not real arguments for or against evaluation. They simply involve different procedures applicable to different stages of the evaluation process.

But there are useful distinctions to be made (see Loomis,1987). Formative evaluation (also called "developmental evaluation") is useful during the planning of new exhibits or when changes to older exhibits are being considered, such as improved instructions, lighting, adding peripheral systems like computers to enhance exhibits, and so on. Summative evaluation (or "post-design" evaluation) applies to exhibit impact following installation and has broader purposes: to identify what might be used (or avoided) in future installations, to identify post-installation improvements, or to identify factors such as relative cost, unexpected behavioral or attitudinal effects (goal-free evaluation), long term effects on behavior/interests/attitudes after leaving the museum, and so on (Screven, 1976, 1986; Shettel, 1973). Other assessment methods are applicable during predesign stages such as audience surveys, analyses of entering knowledge and attitudes, marketing research, and feasibility studies.

Formative Evaluation:
Pretesting Alternative Approaches

During new exhibit planning, layouts, media, labels, models, objects, etc. are chosen usually because the exhibit team assumes they will be more effective than alternatives in attracting viewer attention and interest, conveying information, and so on. These assumptions often are based on intuitive ideas generated by experience, preconceptions about target audiences, or by popular "theories" about visitor characteristics. Often such intuitions and "theories" prove wrong. Clearly, the likelihood of post-installation problems and associated costs is reduced with information on how viewers are likely to respond to prospective exhibit approaches. The same may be said when making post-installation changes. Formative evaluation can be an economical and practical way of obtaining this kind of information.

Formative evaluation involves pretesting design concepts, text/ graphics elements, placement, etc. during exhibit development or during efforts to improve existing exhibit components. It employs low cost, quickly made versions of the most important labels, photos, objects, layouts, and instruction panels which are tested with small samples of target audiences and adjusted before proceeding to later stages. For more details concerning the process, the reader should see the papers of Griggs (1981, 1983), Jarrett (1986), Screven (1976, 1986), and Shettel (1968, 1973).

Misconception: *Formative evaluation requires realistic and to scale prototype displays requiring weeks or months of preparation prior to testing.*

The prospect of expensive preparations discourages pretesting exhibits. It not only is costly, but also psychologically unacceptable. Many problems surface in the first few hours or days of testing. If months of effort has been invested, poor results can be devastating to the morale of a professional exhibit team. While not all questions can be evaluated with simple mockups, they provide valuable early information on communication and motivation issues.

Early formative testing. Early testing uses quickly made prototypes requiring only a few hours or days to develop. Such testing helps identify inappropriate assumptions about visitor knowledge or attitudes, misconceptions that can influence how visitors will interpret exhibit content, the ability of proposed exhibit content and story lines to "connect" to existing visitor knowledge and experience, comprehension of test samples, photographs, or diagrams, the effectiveness of instructions, headlines, questions, artifacts, and so on. At early stages, mock-ups are hand-made or computer generated, employ photos or objects rather than actual objects, etc., and are tested with small random samples (10 to 25) of representative visitors. Materials are set up on a table in a separate

room, laid out in an exhibit area, sequenced in loose-leaf notebook, pasted on display windows or placed on easels, to name only a few of the many possibilities for simulating the formats and content of exhibits. Observations may be made with or without visitor awareness of what is going on (see "cued" and "noncued" testing below).

Visually, mock-ups appear very different from final installations, so there is understandable skepticism that crude early mockups say much about public responses to full scale exhibits in real architectural spaces. However, research has indicated that responses to final installations are remarkably similar to mocked-up versions (Shettel, 1969; Griggs & Manning, 1983).

Formative testing at later stages. Testing at later stages identifies the impact of more developed versions of text, layouts, lighting, neighboring displays and the often subtle roles of social influences (e.g., MacManus, 1986), advance organizers, personality variables, architectural features, and so on. While early mockup testing identifies potential problems in communicating exhibit content (see cued testing below) and behavior at individual display units (e.g., attention, holding power), evaluation at later stages usually involves broader behaviors, disorientation and overall motivational effectiveness (time spent, interaction with and use of exhibit components, social facilitation, sensory overload, etc.). Some effects may not be observable until the exhibition has been opened to the public. At advanced stages, selected groups may be invited to see advanced exhibit models, prototype displays, exhibit areas under construction, and even to preview the exhibition prior to public opening.

Major problems uncovered at this stage are little use if corrective actions are too costly to make. Given earlier formative evaluations, however, major problems with content at later stages that confuse or reinforce misconceptions are unlikely. Most problems will involve motivational factors and spatial and conceptual disorientation arising from text placement, topic headings, lighting, unexpected social and psychological factors, and visual competition among exhibits. These usually can be corrected at very low cost if they have been anticipated (as they should be) and "built-in" to the exhibit design so changes can be made (e.g., snap-on text/graphics panels, etc.). Whenever possible, difficult to change components should be made in nonpermanent forms until adequate testing is possible.

After an exhibit opens, fine tuning of signage and behavioral (motivational) effectiveness, is recommended because all exhibits are in place and actual audiences are in attendance (see Figure 1). Post-installation adjustment costs are minimized if those components most often in need of adjustments (text, graphics, headlines, and orientation devices) have been designed to allow easy changes (see Screven, 1986). Some museums allow up to 10% of exhibition budgets for such post-installation adjustments.

Misconception: *Formative evaluation is a scientific process similar to research studies in other scientific disciplines.*

Formative evaluation is not something you do to discover fundamental relationships among exhibit or visitor variables (this requires careful controls, larger samples, etc.). Formative evaluation provides immediate answers about specific exhibit components or approaches and usually identifies those not likely to work without adjustments. Like summative evaluation, formative evaluation usually does not provide highly reliable information about design features that will work better than other features. . .mainly information about what may not work. It may not even tell you what to do to fix it! Many factors and interactions among these may contribute to visitor usage and learning in museum settings. Thus, usage of interactive "flip questions" (Screven, 1987) depends on the questions being asked, wording, graphics, placement of the information needed to answer the question, placement of the flip label, and its mechanical design. It is almost always prudent and cost-effective to pretest mockups of such complex components to determine which are not likely to achieve the results you desire. But, scientific studies to establish which features are most important are unrealistic. Time and money usually do not allow it.

Misconception: *Formative evaluation is something you undertake until you find something that works.*

Formative evaluation is not a trial-and-error process in which you try out almost any idea and let visitor reactions tell you whether you need to go to something else. Theoretically you could do this, but obviously it is impractical. In reality, you employ formative evaluation to test a limited number of important questions that arise during the design process. You cannot test everything – or even most things. There isn't enough time! If so, then how do you know which objects, labels, graphics, placements, etc. should be pretested?

Formative evaluation is not a substitute for thinking. With or without evaluation, exhibit planning requires some kind of theoretical framework that allows planners to make fairly good predictions about how something will work (Miles, 1988). You need as much knowledge as you can get about who the visitor is, what they like, know and don't know, about how they learn, how they feel, and all the rest. The guts of most successful exhibit designs derive from such knowledge. For example, you may know from published literature and your own experience that visitors pay attention to things that connect to their own experiences. So you select objects, themes, and text that you think will connect to these experiences and communicate your message. Given a good framework about your audiences, museums as teaching-learning environments, and sensitivity to goals, your approach to an exhibit can be generally correct and yet fail after installation!

There are weak links in this complex design process. Specific text can have varying effects; objects and displays can be misinterpreted; there are uncertainties about how media will interact with given materials, as well as lighting, locations, heights, relative sizes, the effects of nearby displays, and many other things. All this makes it difficult to predict important effects--even when your underlying strategies are basically correct! Formative evaluation helps adjust your underlying knowledge and experience at least to some of the unpredictable realities of the open exhibit environment.

Misconception: *Formative evaluation needs to be conducted by highly trained professional evaluators and psychologists.*
When beginning formative evaluation, onsite training by persons experienced in goal analysis, audience analysis, human behavior and learning in public spaces, and formative evaluation is strongly recommended, perhaps with periodic professional followup. But, the bulk of the formative testing process can (and should) be carried out by museum design teams on a day to day basis. Except for initial training and, if needed, periodic outside help, formative testing need not seriously disrupt other routines. Time and costs often are made up by the concrete information obtained about audience responses which reduces guesswork, arguments and costly changes at later stages. The need for a theoretical framework discussed earlier requires someone with a good grasp of human behavior, learning and probably educational psychology. This currently is lacking in many museums which may require new personnel requirements, outside training for selected staff, inhouse training, or the periodic use of specialists.

Misconception: *Formative evaluation requires less need for pre-design planning activities.*
Figure 1 shows the time ideally applicable to different stages of exhibit development. As may be seen, predesign planning involves the largest portion of time. Predesign methods ("front-end analysis") can be different from those applicable during design stages (Miles, 1988; Screven, 1986). Goal-free, "naturalistic" and ethnographic studies (Wolf, 1980) are often very useful here, along with other methods. Predesign planning begins with general notions about exhibit themes and the major audiences to be served, followed by careful definition and prioritizing of goals (using formal, informal, and goal-free observations), and translating goals into measurable form (called "objectives"). Concept maps/exhibit briefs evolve during front-end analysis. The planning stage is followed by the design stage, where formative evaluation methods are most relevant. This is followed by fabrication/installation, summative evaluation of the overall exhibit and, if needed, adjustments to poorly working, easily changed components (headings, text, etc.) again employing formative evaluation of improvements before final installation.

Figure 2 shows the planning and formative stages in more detail. Note that the predesign stage includes two major classes of information:

- Information about prospective audiences (misconceptions, intrinsic motivations, existing knowledge or experience that may serve as starting points for making "contact" with them, time constraints, learning styles, etc.) and
- Information about the exhibit's topic from which exhibit content and "messages" (teaching points) will be derived. "Messages" not only may include facts and verbal information, but also "feelings," a desire to find out more, learning that learning is "fun," learning to learn, etc. Questions of content involves testing prospective materials with subject-matter experts (not visitors) for errors, distortions, etc.

Misconception: *Formative evaluation applies to the communication impact of exhibit elements, not to their behavioral impact, which must await final installation.*

Both the communicative (learning) effectiveness of exhibit components and some aspects of visitor attention, time spent, cooperation and other behaviors can be tested during early stages. But different methods are needed.

(a) Cued Testing: Do exhibit elements communicate?

When visitors are sufficiently motivated to read or attend to message elements of a display, do they comprehend the desired message? If so, these elements are said to communicate. Cued visitors are aware they are to be observed and questioned. Randomly sampled visitors are asked to look at (read, explain, etc.) text, diagrams, objects, mock-up displays, etc., or react to story line of exhibit script. Knowing they will be questioned, cued visitors typically give greater effort and attention to the exhibit materials. Thus, the results reflect the ability of the materials to communicate under motivated conditions. If the majority of 15 or 20 visitors are confused or miss key points, the content probably needs improvement. Cued testing also allows visitors to express their interpretation of things, topics, ideas, etc. which helps to identify inappropriate teaching points and organization of materials. If cued visitors do not comprehend display materials or show deficiencies that can affect their ability to understand the display, the problem probably lies with the teaching elements of the exhibit – vocabulary, clarity, organization, misconceptions, poor connections to visitor knowledge or experience – which can then be modified. Cued testing does not reflect how visitors will respond who are unaware they may be questioned.

(b) Noncued Testing: Do visitors voluntarily attend and cooperate?

This involves the "motivational" impact of exhibit elements and includes attraction (approaching or stopping or noticing an exhibit component), time spent (total time), holding power (time spent in

relation to minimum time needed), interaction (manipulate in appropriate ways), cooperation (follow instructions, utilize exhibit information before responding), reading, exhibit linked socialization (talk about exhibit content, point), and emotionally involvement (smile, frown, fidget). Comparing the performance of randomly chosen visitors after leaving an exhibit with similar visitor sample before seeing the exhibit is one way of evaluating an exhibit's impact on learning. But, this does not mean it is communicating efficiently – visitors may be learning because its content is highly motivating – i.e., draws and holds attention. Therefore, learning by noncued visitors reflects both communication and motivational influences. Learning has occurred, but not necessarily because it communicates well. (For example, a poorly written book can produce "learning" if the reader has to qualify for a job!)

The challenge is to maximize the efficiency by which an exhibition communicates its major teaching points (thus reducing the effort and time required) and provide motivation that encourages adequate time and effort (assuring that its content can be communicated. `

Summary

In summary, formative evaluation:

• usually can be conducted during very early stages of exhibit development with simple mockups;
• is not formal scientific research of variables that tests hypotheses about exhibit/visitor characteristics. It provides practical feedback about behavioral or learning impact of specific materials during the design stage, not why given effects do or do not occur;
• can be carried out by museum exhibit teams after training;
• does not substitute for thinking. Designing educational exhibits assumes a theoretical framework -- an understanding of behavior and learning in public areas. Formative evaluation reduces the likelihood of errors arising from many unknown influences operating in open exhibit environments;
• assumes serious attention has been given to "front-end" planning (the prospective audience, analysis of the exhibit's messages, goal setting and organization of teaching points);
• is applicable to both communication effectiveness and the impact on visitor behavior of prospective displays during early and later stages of exhibit design;
• replaces design decisions based on hypothetical audiences with a more realistic view of real audience capabilities and limitations;
• avoids the cost and demoralizing effects of failures revealed by later summative evaluations.

References

Griggs, S. A., and Manning, J. (1983). The predictive validity of formative evaluations of exhibits. Museum Studies Journal, 1 (2), 31-41.

Jarrett, J. E. (1986). Learning from developmental testing of exhibits. Curator, 29 (4), 295-306.

Loomis, R. (1987). Museum visitor evaluation: New tool for museum management. Nashville, TN: American Association for State and Local History.

McManus, P. (1987). It's the company you keep: The social determination of learning-related behaviors in a science museum. International Journal of Museum Management and Curatorship, 263-270.

Miles, R. S., Alt, M. B., Gosling, D. C., Lewis, B. N., and Tout, A. F. (1987). The Design of Educational Exhibits, 2nd Edition. London: Allen and Unwin Publishers.

Miles, R. S. (1988). Exhibit evaluation in the British Museum (Natural History). ILVS Review: A Journal of Visitor Behavior, 1 (1).

Screven, C. G. (1976). Exhibit evaluation: A goal-referenced approach. Curator, 19 (4), 271-290.

Screven, C. G. (1986b). Exhibitions and information centers: Principles and approaches. Curator, 29 (2), 109-137.

Screven, C. G. and Gessner, K. (Eds.) (1988). ILVS bibliography and abstracts, 2nd Edition, 1 (2). Milwaukee WI: International Laboratory for Visitor Studies, Department of Psychology, Garland 138, University of Wisconsin-Milwaukee. 88 pp.

Shettel, H., Butcher, M., et al. (1968). Strategies for determining exhibit effectiveness. Final Report, AIR E95-4168FR. Washington, D.C.: American Institutes for Research. (Also available from ILVS).

Shettel, H. (1973). Exhibits: Art form or educational medium? Museum News, 52, 32-41.

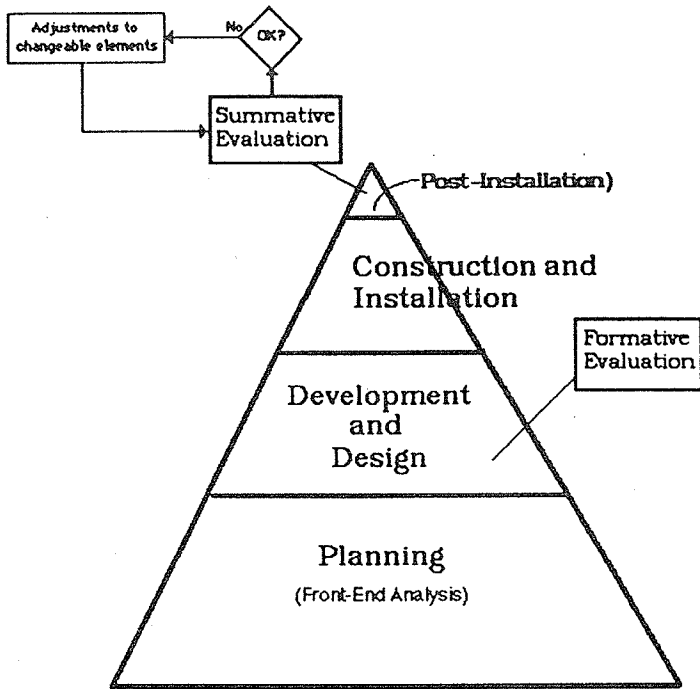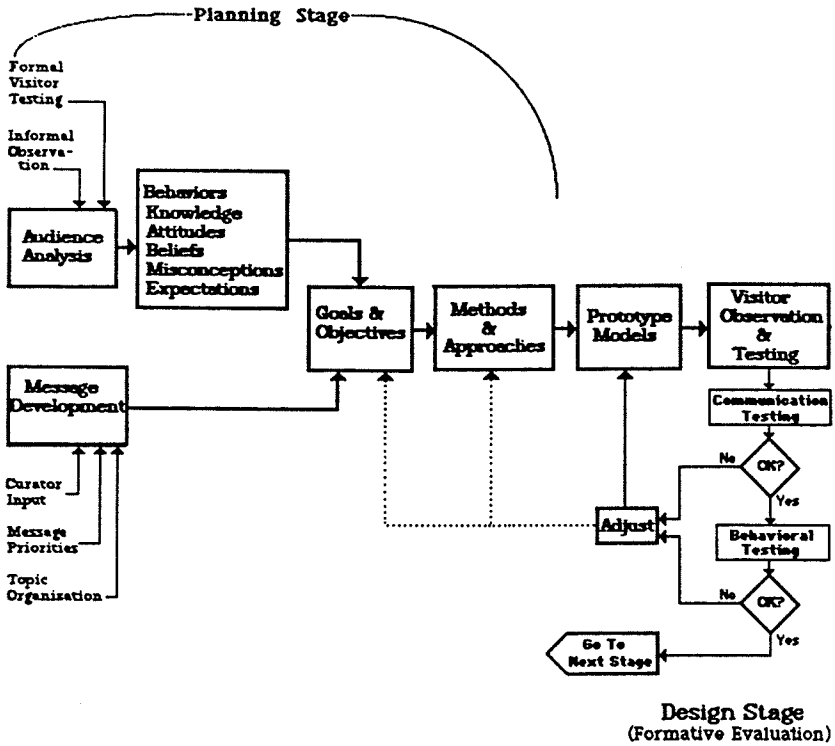Wolf, R. L. (1980). A naturalistic view of evaluation. Museum News, 58 (1), 39-45.

Figure 1

Figure 2. PLANNING AND EVALUATION STAGES IN
EXHIBITION DEVELOPMENT
(Adapted from Screven, 1988. C 1987 C. G. Screven)