

## Classification of Exhibit Evaluation: How Deep Should Occam's Razor Cut?

Stephen Bitgood  
Jacksonville State University

Roger Miles (1994) in this issue has appealed to the principle of Occam's Razor in determining the number of exhibit evaluation types. This principle argues for parsimony, i.e., the fewest number of assumptions should be made when explaining something. All other things being equal, this is an admirable principle. But, Occam's Razor should not cut so deep as to remove the heart of important differences in classification. Occam's Razor should be used only with extreme surgical care lest it do serious damage.

I agree with Miles' argument in this issue that a classification system of exhibit evaluation should meet at least three criteria: (1) it should order our knowledge; (2) it should promote clear thought; and (3) it should be parsimonious. I also agree that evaluation of exhibitions is never finished — it should be applied during the entire life of an exhibition. I disagree, however, on how the different types of evaluation should be defined and thus how our knowledge should be ordered to promote clarity of thought and communication.

Screven (1990) and Miles (1993; 1994) have proposed different classification systems for exhibit design and evaluation. Screven proposes a system with four types of evaluation, while Miles argues that only three types are necessary. Screven distinguishes among five stages of exhibition development (planning, design, construction/installation, occupancy, and remedial), while Miles includes only three (plan-

ning, design, and post-installation). Miles argues that evaluation types should be based on *when* evaluation occurs and *what* is being evaluated; Screven describes additional dimensions including the purpose of evaluation.

Bitgood and Shettel (1994) recently defended the use of remedial evaluation arguing that purpose should be used as a defining characteristic of evaluation type. Purpose was defended on two grounds: (1) the distinction between formative and summative evaluation has traditionally included purpose; and (2) purpose is an important aspect of evaluation. Although it was not explicitly stated, Bitgood and Shettel adapted Miles three stage model (planning, design, post-installation) rather than Screven's five stage model (planning, design, construction/installation, occupancy, remedial). The purposes of this paper are: (1) to argue further for a classification system which includes four types of evaluation, and (2) to expand our earlier rationale (Bitgood & Shettel, 1994) for such a system.

Perhaps it is an oversimplification to think that evaluation types can be defined by only two dimensions (whether these dimensions are when/what, or when/purpose). It seems to be that if we are to order our knowledge and promote clear thought, classification should be based on multiple dimensions: *when* evaluation occurs, *why* evaluation is being conducted (purpose), *how* evaluation is undertaken, *what* is being evaluated, and *who* or how many visitors are selected for evaluation. Most of these dimensions seem to be imbedded in Screven's (1990) article.

### AN OVERVIEW OF TYPES OF EVALUATION

Each type of evaluation is described below in terms of five dimensions. Figure 1 provides a summary of this analysis.

Figure 1

TYPE OF EVALUATION	WHEN	WHY	HOW	WHAT	WHO
FRONT-END EVALUATION	Planning stage	Gather information on visitor knowledge, misconceptions, and interests	Administer survey to potential users	Plans Exhibit ideas Concepts	Moderate-sized sample
FORMATIVE EVALUATION	Design stage	Improve exhibit elements before production	Trial test makeshift exhibits using an iterative process	Mock-ups and other elements of individual displays	Small sample
REMEDIAL EVALUATION	Post-installation	Improve exhibit elements after production	Trial test completed exhibits using an iterative process	Various elements of individual and groups of exhibits	Small sample
SUMMATIVE EVALUATION	Post-installation	Assess the overall impact of installed exhibits	Comprehensive study of overall exhibition performance	Entire installed exhibit	Large sample

### Front-end Evaluation

1. *When is it done?* Front-end evaluation is conducted during the planning stage. The classification systems of both Screven and Miles agree on this dimension. Miles has suggested that the planning stage ends with a written brief (Miles, 1994, p. 4).
2. *Why is it done?* Front-end evaluation is undertaken to: (a) gather information on user knowledge, experience, misconceptions, interests and attitudes; (b) improve and refine teaching points, select media, design strategies; (c) identify sources of confusion, resistance to exhibit material; (d) decide on priorities for teaching points and goals; and/or (e) help select media and design strategies (identify media preferences, visitor learning styles).
3. *How is it done?* Information is obtained from potential users by means of interviews, questionnaires, and/or focus groups. Survey items can be open-ended (e.g., "What comes to mind when you think about Asian art?") or require more quantitative answers such as in a rating scale (e.g., "Rate your interest in Asian art from '1' - 'Very little interest' to '7' - 'Extremely interested'").
4. *What is being evaluated?* The exhibition plans are the focus of front-end evaluation. Thus, content of the surveys are centered on the subject matter of the to-be-developed exhibition.
5. *Who is being studied?* Sampling is conducted on a moderate-sized group of potential users. Although it is difficult to fix an exact number for a survey sample, at least 25 individuals per sub-group is advisable (Borun, 1992). For example, if the results are going to be divided into adult and family groups, then at least 25 adult groups and 25 family groups should be selected.

### Formative Evaluation

1. *When is it done?* Formative evaluation is conducted during the design stage of an exhibition project. "It begins with temporary versions of design ideas and proceeds to the evaluation of increasingly refined and integrated displays" (Screven, 1990; p. 41). The design stage ends when production drawings are ordered.
2. *Why is it done?* Formative evaluation is undertaken primarily to improve individual exhibit displays that are usually selected because they have the greatest potential for communication problems. The purpose is to obtain "information about visitor reactions to temporary versions of the most important panels, formats, text, displays, etc. in terms of both their ability to generate and focus visitor attention and effort, and their ability to 'deliver' (communicate) their message." (Screven, 1990; p. 41). It provides information about what visitors do or feel or comprehend when confronted with instructions, labels, objects, layouts, topics.
3. *How is it done?* Formative evaluation involves trial testing exhibit displays using an iterative process (i.e.,

testing, modifying, retesting, etc.) until they meet their performance goals.

4. *What is being evaluated?* Mock-ups, prototypes, working models, makeshift displays, and/or temporary exhibits are assessed during formative evaluation. Emphasis is on improving individual displays and consequently within-display variables (e.g., wording of labels) related to understanding and comprehension of the exhibit's messages are the usual focus.
5. *Who is being studied?* Small samples of potential visitors are selected in formative evaluation. If it becomes obvious after testing five visitors that the directions for using the exhibit display are confusing, one would immediately stop collecting data, make a change in the directions, and then resume testing the modified element.

### Remedial Evaluation

1. *When is it done?* Remedial evaluation can be conducted any time after an exhibition is open to the public.
2. *Why is it done?* It is undertaken to find ways to improve some or all elements of an installed exhibition. Screven (1990) suggests that it can deal with problems created when all of the individual elements and displays of an exhibition are actually placed together on the floor and are subject to other variables such as architecture (e.g., doors, windows, lighting).
3. *How is it done?* Remedial evaluation is accomplished by trial testing exhibit and non-exhibit factors using an iterative process similar to formative evaluation. Thus, some aspect of the installed exhibition is assessed using visitor reaction, improvements are attempted (e.g., adding an orientation panel), and the impact of the modified exhibition is re-assessed.
4. *What is being evaluated?* Remedial evaluation assesses exhibit changes such as label content, lighting, sight lines, choice points, and furniture (location, additions). The focus is usually on the effectiveness of individual elements in the context of surrounding objects, architecture, etc.
5. *Who is being studied?* Small to moderate samples of visitors are used to assess the impact of exhibition elements. In some cases, only a few users are necessary to determine if an exhibition element is performing poorly. However, for assessing the impact of other elements, a larger sample may be necessary. For example, if one is concerned about the attracting power of a newly designed orientation label, a reasonable size sample may be necessary.

### Summative Evaluation

1. *When is it done?* As with remedial evaluation, summative evaluation can be conducted anytime after installation.
2. *Why is it done?* The purpose of summative evaluation is

to assess the overall impact of installed exhibitions. It is intended to provide information on how the exhibit is working overall, how people use it, what they learn from it, or how they are changed by it. (Screven, 1990; p. 52). It is used to provide reports, to plan future exhibitions, suggest research, and can also be used to point to possible problems for remedial evaluation if such evaluation has not already taken place. The kinds of information sought include: educational, behavioral, affective impact of the overall exhibition and its surrounding space; what might be useful in planning future exhibits; cost comparisons with similar exhibits; identifying unintended effects not part of original goals; broad impact of exhibit on larger community, museum attendance, etc.; research data that can test theories of informal learning (Screven, 1990).

3. *How is it done?* Summative evaluation involves a comprehensive study of overall exhibition performance. It measures attraction power, holding power, and teaching power of the entire exhibition. It might include tracking visitors through the exhibition, survey procedures, focus group studies, cued testing at individual exhibits, and rating scales for exhibit effectiveness (Screven, 1990; p. 53). Unlike remedial evaluation, however, it does not include an iterative process (test, modify, re-test).
4. *What is being evaluated?* The performance of the exhibition as a whole is the focus of summative evaluation, rather than individual elements as is the case of remedial evaluation.
5. *Who is being studied?* A large sample of pre- and post-visitors is necessary in order to make inferences about the overall success of an exhibition. Statistical analyses of results is often an important element of summative studies.

## ISSUES IN CLASSIFICATION

### Remedial Evaluation

Miles (1993; 1994) has at least two arguments against the use of remedial evaluation. First, he argues that one cannot define evaluation based on its purpose. However, others appear to have little problem describing the purpose of evaluation. It is not difficult to make a clear statement about why one is undertaking evaluation. In fact, failing to make such a statement about intention can lead to considerable confusion.

A second argument that Miles makes against the use of remedial evaluation is that it creates overlapping categories between remedial and summative evaluation. He points out that an evaluator's actions might be both summative and remedial at the same time. However, an evaluation type should not be defined by an evaluator's action or exclusively by an evaluator's purpose. Rather, it should be defined by multiple dimensions. As indicated by Figure 1, remedial and

summative evaluation differ along several dimensions: why it is done, how it is done, what is being evaluated, and how many individuals are being studied.

### Number of Stages

The three stages of Miles (planning, design, and post-installation) appear to be sufficient to order our knowledge. As Miles points out, the distinctions among Screven's last three stages are not always clear.

### Defining Dimensions of Evaluation

Miles' classification model limits defining characteristics to when evaluation occurs and *what* is being evaluated. This restriction seems to ignore other important differences among evaluation types. The *why*, *how*, and *who* of evaluation are ignored. This article has argued that these dimensions should also be considered in a classification system.

## SUMMARY AND CONCLUSION

The current article: (1) offers a rationale for Screven's (1990) four-type evaluation model based on multiple dimensions (when, why, how, what, and who); (2) accepts Miles (1993; 1994) three-stage model of evaluation; and (3) addresses Miles' criticisms of the use of remedial evaluation as one of the evaluation types.

Remedial evaluation differs from both summative and formative evaluation on several dimensions and therefore should be considered a separate type of evaluation. I believe that the principle of Occam's Razor or parsimony cannot be used as a rationale to cut remedial evaluation from the list of evaluation types since it interferes with clear thought and does not sufficiently order our knowledge of the evaluation process. Remember Whitehead's Rule: "Seek simplicity, and distrust it."

## References

- Bitgood, S., & Shettel, H. (1994). The classification of evaluation: A rationale for remedial evaluation. *Visitor Behavior*, 9(1), 4-8.
- Borun, M. (1992). Exhibit Evaluation Workshop. Baltimore, MD.
- Miles, R. (1993). Grasping the Greased Pig: Evaluation of Educational Exhibits. In S. Bicknell & G. Farmelo (eds.) *Museum Visitor Studies in the 1990s*. London: Science Museum of London, Pp. 24-33.
- Miles, R. (1994). Let's Hear It for Mr. Occam: A Reply to Bitgood & Shettel on Remedial Evaluation. *Visitor Behavior*, 9(3), 4-7.
- Screven, C.G. (1990). Uses of evaluation, before, during, and after exhibit evaluation. *ILVS Review*, 1(2), 36-66.