Let's Hear It for Mr. Occam: A Reply to Bitgood & Shettel on Remedial Evaluation

Roger Miles
The Natural History Museum
London

In a paper succinctly summarizing the history of the classification of exhibit evaluation, Bitgood and Shettel (1994) express their disagreement with my classification (Miles, 1993), which is shown in Figure 1.

This classification is based on the *when* (in relation to design and production) and the *what is evaluated*, of evaluation. Bitgood and Shettel, on the other hand, uphold Screven's (1990) classification (Figure 2), which differs in including a fourth term, remedial evaluation, and in using the *purpose* of evaluation as one of its defining criteria. The aim of this paper is to defend my 1993 proposal by showing that the concept of remedial evaluation is unnecessary.

What is Remedial Evaluation?

Screven (1990) provides a description rather than a formal definition of remedial evaluation. From this we glean that:

(1) It takes place at the remedial stage of exhibition development, which follows the occupancy stage and summative evaluation. This is the criterion of *when*.
(2) It "tests soft mockups and prototypes" (Screven, 1990; p. 54). This is the criterion of *what*. However, remedial evaluation adjustments, "usually involve changes or additions around the exhibits rather than, *in the individual exhibits themselves*. . ." (p. 55).
(3) It has the task of reducing "the disruptive effects of variables introduced by the macro-exhibition environment" (Screven, 1990; p. 54). This is the criterion of *purpose*. These "occupancy variables" may be of a physiological (e.g. fatigue), architectural (e.g. competing exhibits), social (e.g. competition) or psychological (e.g. information overload) nature.

Let us now examine in more detail the use of these criteria in Screven's classification.

When

The three stages of exhibition development in my scheme are well delimited one from the other. Thus the end of the *Before* stage signals the completion of preplanning and authorization; the mission is defined: "This will normally include a synopsis of the proposed intellectual content, thereby clarifying what the exhibition is to be about. It must also consider who the exhibition is to be aimed at, what the designers are trying to achieve, and what sort of resources are available" (Miles et al, 1988; p. 11). These statements of purpose are contained in the written brief (or program) which, once authorized, is the physical evidence that the *Before* stage is complete and the *During* stage can start. The During stage ends with the opening of the exhibition.

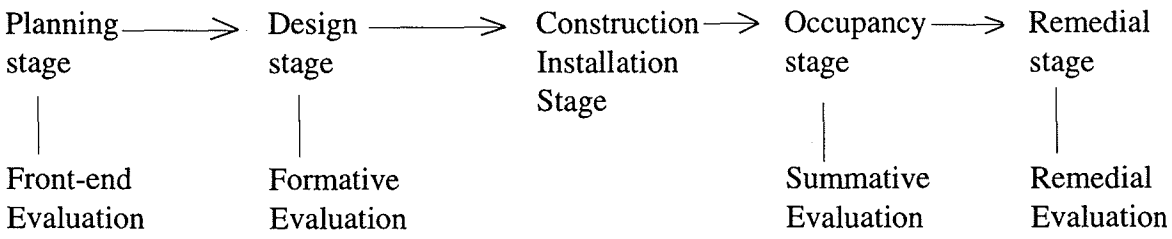
No such clear distinction separates Screven's *Occupancy* and *Remedial* stages. The two would appear to merge, and this has consequences for the type of evaluation that is supposed to take place at each stage. If I am correct in assuming that his comment, "Redefining the exhibit development period to include six to twelve weeks following its public opening is not un-doable. . ." (p. 55), alludes to remedial evaluation, it would seem inevitable that, on the criterion of *when*, it overlaps with summative evaluation. There is simply insufficient time to debug and summatively evaluate the exhibition before remedial evaluation starts.

My experience leads me to say that, at least with a major educational exhibition, the first six to twelve weeks after opening are devoted first to letting the team recuperate; and secondly to making the exhibition work as planned, often in

Figure 1
Miles (1993) Classification of Evaluation

	Before	During	After
Plans	Front-end		
Mock-ups		Formative	
Opened exhibits			Summative

Figure 2
Screven's (1990) Stages of Exhibition Development



fairly crude "mechanical" terms. The latter does not require formal evaluation, whether styled summative or remedial. Inter-ocular evaluation is enough: the problems hit you between the eyes, and there is no need for a sophisticated study to tell you what to do. Because there are urgent and obvious problems to attend to, there is no point in trying to unearth and illuminate subtle problems. These may safely be left to summative evaluation, which will come along in due course when everyone is ready (in other words, evaluation is something that has to be managed). We should not lose sight of Melton's (1935; p. 85) methodological point, "whenever a new permanent display is to be used, a period of at least several months should elapse before starting the [experimental] investigation."

What

Remedial evaluation makes use of mockups and prototypes to elucidate *between-exhibit* variables. Remedial evaluation, is "likely to focus on physical and architectural features like lighting, placement of thematic headlines, banners, entrance-exit designs, sight-lines, and choice points and psychological factors like disorientation, crowds, thematic layout, information overload, fatigue, social activity, and so on" (Screven, 1990; p. 55). However, I fail to see why *all* of this involves testing "soft mockups and prototypes". Some of it may, but other aspects (e.g. lighting, sight-lines) would appear more to involve moving the exhibition furniture around.

A specific example given by Screven (1990; Table 4) as a starting point for correcting occupancy problems is: "add seating among and between exhibit areas" (other examples are more cognitively oriented). But why should such changes not be the outcome of summative evaluation? After all, it seems perfectly possible to move a bench into an exhibition without calling it a prototype and dignifying the action as remedial evaluation. The justification would appear to be that summative evaluation is something done, "only to determine the overall success of the completed exhibit" (Bitgood and Shettel, 1994, p. 5). But this self-denying ordinance is just that: self-denying. It runs counter to experience (e.g., of bench moving, or of breaking text into small paragraphs), in which action flows directly from evaluation without the need to test mockups or prototypes.

Thus, as described by Screven, remedial evaluation embodies elements of formative (use of mock-ups) and summative (occupancy context) evaluation; but this duality is not its defining character, which "is clearly anchored to the dimensions of *when* it occurs and its *purpose*" (Bitgood and Shettel, 1994; p. 6). Thus the seeds of confusion are sown. We have seen that their development is not arrested by the critierion of *when*, because it conflates remedial and summative evaluation. We must now look at *purpose*.

Why Purpose is a Poor Criterion

My main reason for asserting that purpose cannot be used to discriminate between remedial and summative evaluation is that, as a creative human activity, evaluation simply cannot be constrained in this way. One of Bitgood and Shettel's examples will serve to make this point. Referring to Scriven's (1967) definitions of formative and summative evaluation they write:

> Scriven's terms are commonly used in formal education today. For example, in higher education course evaluation instruments are often used as feedback to faculty for self-improvments (formative evaluation) and/or for making decisions about promotion and merit pay (summative evaluation).

But as managers know, you cannot separate "making decisions about promotion and merit pay" from giving feedback. This is because persons who are not favorably judged want to know, and have every right to ask, what they must do to improve their performance, and thus qualify for promotion and merit pay the next time round. So immediately a distinction based on function breaks down, because the outcome of the evaluation is not known in advance.

Another way of looking at this problem (and of bringing us back to our remedial vs summative problem) is to ask what the evaluator is doing when he or she starts to evaluate an exhibition in say the eleventh week after opening. If the results lead to the redesign, or reconfiguration, of some exhibits it is remedial evaluation, at least according to the

criterion of *purpose* ("improvements"), whether or not mock-ups are used. If they lead to a decision to do nothing, at least for the foreseeable future, then it is presumably summative evaluation, again according to the criterion of *purpose* ("determining success"). We all aspire to 20:20 vision in hindsight, but we would surely prefer our evaluator to be able to put a name, *at the outset*, to what he or she is doing. But apparently this is impossible, because summative evaluation can, at any moment, mutate into remedial evaluation. To err is human, to fail to deal promptly with obvious mistakes is ill-advised (and if you are attempting to innovate, a gift to reactionary critics).

Of course the purpose of remedial evaluation is to remedy mistakes, but so is that of front-end and formative evaluation. However, we also accept that front-end and formative evaluation have a role in *unearthing* mistakes. There is no reason why remedial evaluation should not also be given this role. When it is, it becomes synonymous with summative evaluation. One way of saving the situation might be to define remedial evaluation as post-occupancy evaluation that makes use of mockups and prototypes. However, this leaves a lot of very closely related work, which also has the purpose of making improvements, in the domain of summative evaluation. So this suggestion will not get us out of our predicament.

The root cause of these problems lies in the linear, *life-history* model of exhibition development that is in Figure 2. I suggest we should think in terms of the *life-cycle* of an exhibition.

The Life Cycle Model

The essence of this model (Figure 3) is that, as a result of summative evaluation:

1. Changes may be made to an exhibition without further ado;
2. Exhibits may individually be redesigned or replaced, which cycles us back to the Design stage, which may involve formative evaluation;
3. The entire exhibition may be replaced, which recycles us to the pre-planning or planning stage, and possibly to front-end evaluation.

All this requires are two small "conceptual leaps": (1) that formative evaluation take place in the post-occupancy exhibition environment, and (2) that summative evaluation can directly inspire action to improve exhibit performance.

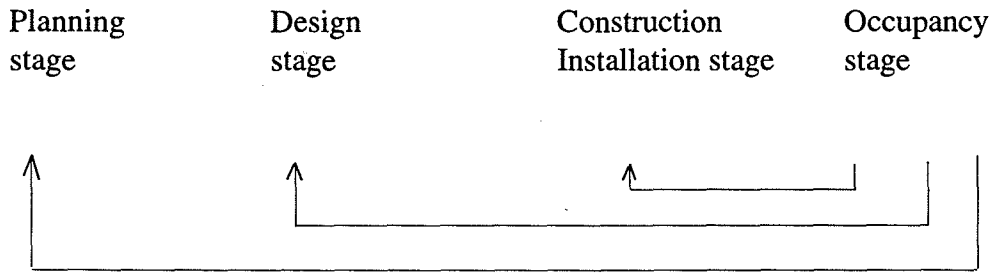
Changes may be made without the benefit of formal evaluation. In most cases I am inclined to regard any informal summative evaluation that takes place, as arising out of critical appraisal. Screven (1990; p. 52) regards critical appraisal as something "conducted by experts before (or after) a summative evaluation with visitors is initiated." However, I prefer not to regard it as a *type* of evaluation, but rather as an informal way of collecting data, that may be employed in front-end, formative or summative evaluation (Miles, 1993; Table 2).

The concept of an exhibition life-cycle is closely related to the view that work on an exhibition is never finished, Screven and I agree on this point. Bitgood and Shettel (1994; p. 8) write:

Miles and Clark (1993) state that evaluation after the exhibit is installed "is generally too late - the money has been spent, and there is none left to correct mistakes." Screven (1990) on the other hand, argues that a percentage of the total budget should be reserved for improvements after the exhibit is installed.

The use of the words, "on the other hand," hints at conflict. In fact, there is nothing incompatible about these quotations; the first purports to be statement of fact, open to empirical investigation, while the second is a statement of policy.

Figure 3
The Cyclical Nature of Exhibition Development
Showing Three Possible Outcomes of Summative Evaluation
Stages after Screven (1990)



So What?

The purpose of classification is to order our knowledge and promote clear thought. I believe that my classification (Figure 1), which is based on the cyclical model of exhibition development, not only cuts through the muddle explored in this paper, but that is also has formal properties which make it superior to the alternative championed by Bitgood and Shettel (Figure 4). These are:

1. It avoids overlapping categories, such as I have shown to be the case with remedial and summative evaluation (e.g. an evaluator's actions might be both summative and remedial at the same time). This is true of all good classifications: instances of the things being classified should be assignable, *unequivocally*, to one, and only one, category in the classification.

2. It does not offend Occam's razor, i.e. the principle of the 14th century English philosopher William of Occam, which states that the fewest possible assumptions are to be made in explaining a thing. A 25 percent reduction in the types of evaluation we need to name and define is not to be sneezed at, in these days of information overload.

References

Bitgood, S., & Shettel, H. (1994). The classification of evaluation: A rationale for remedial evaluation. *Visitor Behavior*, 9(1), 4-8.

Melton, A.W. (1935). *Problems of installation in museums of art.* Washington, D.C.: The American Association of Museums.

Miles, R. (1993). Grasping the greased pig: Evaluation of educational exhibits. In S.Bicknell and G. Farmelo (eds.), *Museum visitor studies in the 90s.* London: Science Museum of London, Pp. 24-33.

Miles, R., Alt, M., Gosling, D., Lewis, B. & Tout, A. (1988). *The design of educational exhibits.* 2nd Ed. London: Unwin Hyman.

Miles, R., & Clarke, G. (1993). Setting off on the right foot: Front-end evaluation. *Environment and Behavior,* 25(6), 698-709.

Screven, C.G. (1990). Uses of evaluation, before, during, and after exhibit design. *ILVS Review*, 1(2), 36-66.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (ed.), *Curriculum Evaluation.* American Educational Research Association Monograph Series on Evaluation, No. 1. Chicago: Rand McNally.

Note: I am grateful to my colleagues Giles Clarke and Katie Edwards for their comments on my first draft of this reply.

Figure 4
Screven's (1990) Classification of Evaluation

	Before	During	After
Improvements	Front-end	Formative	Remedial
Determining Success			Summative

Factoids on VSA

- The Visitor Studies Association increased its membership from 1993 to 1994 by over 100 which represents an increase of over 30 percent.
- About one-fourth of the VSA membership is from non-USA countries including Canada, Australia, England, Germany, Mexico, and many others.
- California is the U. S. state with the most members (26). Other states with significant numbers are New York, Illinois, North Carolina, District of Columbia, Ohio, Pennsylvania, and Michigan.
- VSA has had two presidents since its birth – Harris Shettel and Ridgeley Williams.