

NEW METAPHORS FOR CARRYING OUT EVALUATIONS IN THE SCIENCE MUSEUM SETTING

Mark St. John^{1,2}

Inverness Research Associates

The prevailing metaphor underlying most educational research over the past twenty-five years has been that of the randomized, controlled experiment. Over the last decade this experimental approach has encountered a great deal of criticism, particularly from those working in the domain of informal science education. Those criticisms are well known to the readers of this paper and will not be repeated here. But before completely dismissing it, let us look more closely at the metaphor that underlies much of the experimental research that has taken place in the museum setting.

The experimental research paradigm that seeks statistical significance is perhaps best described as deriving from an agricultural research paradigm. For example, two separate acres of corn are planted, and on the experimental plot the treatment – a new fertilizer – is applied. All of the other variables are held constant. At the end of several months the ears of corn are picked, examined, and weighed. It is important to note that the corn picked from each plot shows great variation – some ears are very small, some are very large. At first inspection there may be no obvious difference between the corn grown on the experimental plot and the corn grown on the control plot. Perhaps even the average weights between the two plots don't differ by too much – maybe 12.5 ounces versus 13 ounces. The question then becomes: does this difference in averages reflect real physical differences or is it an expected fluctuation in the data. Now, this is where statistical methods are used to tell us how likely it is that the average of one plot is in fact different from the average of the other plot.

Note that the differences may be statistically significant, but in fact may be quite minor in terms of their absolute or material significance. Note also that the corn is being compared along one dimension only – that of weight or perhaps size – but that many other qualities of the corn are deliberately being ignored. Finally, note that this is a treatment-effect experiment. The fertilizer is seen as a treatment, and we are trying to eek out a statistically significant correlation that allows us to make inferences about causal relationships. But note again that there is no information here about mechanisms. That is, we do not learn anything about the micro-mechanisms by which the fertilizer may be enhancing the final weight or size of the corn.

Much past educational research basically lies within this paradigm or ascribes to this underlying metaphor. We test classes of students as if they were ears of corn, and we treat instruction much like fertilizer. We often do not understand the mechanisms, and we often ignore other salient and important qualities (e.g. does the corn taste good? Is the child happy about what he learned?).

Exhibits are sometimes seen as the "treatment," and gains in conceptual knowledge as the "effect." The metaphor here is the exhibit as a teaching machine. Pre- and post-tests allow us to determine the effect of interacting with the exhibit without ever knowing the qualities of the interaction that took place. Such experiments by necessity are limited to testing along one or two dimensions (and they must be quantifiable dimensions at that).

This approach is highly limited at best and wrong-minded at worst. Practitioners often reel at the prospect of their exhibits being evaluated in this light.

Perhaps the above scenario is too simplistic and I have simply created a "strawman" to oppose. But the point is not that rigorous experimentation is inappropriate to museum research, but rather try as we might, we often find our thinking caught within this paradigm.

In this short and informal treatise I will suggest that other metaphors (embedded in other paradigms) may be useful to us as a departure point for our thinking about assessment in the informal domain.

Metaphors

A metaphor (see note 1) is a very useful mental device for helping us gain insight into an area by juxtaposing language and concepts that are intuitively familiar in one area with a new area. One gains insight into relationships previously unseen by using the "lens" or cognitive structures that are supplied by the metaphor.

Every student in physics knows the utility of thinking about electricity using the more familiar metaphor of water flow. Voltage is compared to pressure drop; current is compared to water flow; resistance is compared to blockages in the fluid flow.

Within the field of evaluation, I would argue that we need new underlying metaphors to shape the approaches and techniques we use. New metaphors may help us escape from the rather deep positivistic rut in which our thinking currently runs when we begin to examine the whole question of assessment and evaluation. New metaphors may also help give credibility to asking and pursuing different evaluation questions. They may also give us a range of new tools and new assessment procedures that are compatible with the nature of the informal learning domain of science museums.

In the early 1980's I worked with Nick Smith, of Northwest Labs, on a project which explored alternative metaphors for evaluation. Almost every discipline, field, and endeavor has its own approach for gathering information about what it is doing, for reflecting on its own activities, and for making

judgments about the quality of its own activities. Poets, architects, congressmen, writers, artists, system design engineers, geographers, and philosophers all have their own approaches and techniques for carrying out evaluative activities within their own fields. In the Research on Evaluation Project we examined what practitioners in other fields did – what questions they asked, and what approaches they took to answering them. We looked for the metaphors that underlay their approaches and we looked for ways that they might be applied to the educational domain. In the rest of this paper, I am going to draw on this work to suggest some alternative metaphors for doing evaluation in the museum setting.

The reason I am focusing on this perhaps somewhat abstract set of notions is that I am convinced that unless we go beyond the bounds of the agricultural paradigm in which most of our thinking is centered, we are unlikely to come up with very much that is new and interesting in this project. I am not suggesting that these are the only metaphors or the most productive ones to guide the research efforts that we will undertake. I am simply laying out some alternative metaphors, alternative conceptions, and alternative techniques in order to provoke our thinking.

Architecture

Like an exhibit designer, an architect is concerned with the process of design, the process of inventing physical solutions to physical/social problems. Successful design requires that the architect (exhibit designer) operate both in a linear, analytical fashion and in a more intuitive, holistic mode. The architect (and designer) must possess a wide range of technical skills and knowledge and integrate the methods of many varied disciplines.

Both architect and exhibit designer must understand the multi-dimensional context in which the structure (exhibit) is to function. They must understand the needs of the multiple audiences they are serving, they must operate within constrained resources, their products must fit the physical site, and they must observe strong social and legal constraints.

Both exhibit designer and architect must draw upon their analysis of the context and upon their creative vision as well in order to produce a form that will satisfy as many of the constraints and goals as possible. The form they create is the realization of the design process. It is the solution which it is hoped will fulfill those dimensions of the needs most demanding of attention.

How then do architects evaluate the forms they produce? In architecture, the key concept for evaluation is known as "fit." The fit is the degree of congruence between the form and the context. It is a measure of acceptability. Out of the evaluation of the fit, judgments are made about how well the design has met the demands of the context. Standards of "consistency," "integrity" and "aesthetics" are applied. Poor designs lead to such judgments as "arbitrary, obsolete, incongruous, or dysfunctional." Successful forms, ones that fit the

context well, are judged to be "efficient, compatible, energy saving, and proportional."

Perhaps there is something in the notion of "fit" that can guide evaluators in the museum setting. Exhibits are nothing else if not designs. And the museum is a complex and multi-dimensional context for those designs. The metaphor of architecture allows us to escape the notion of exhibit as a teaching machine. Rather, this metaphor conceives of the exhibit as a design that tries to satisfy many simultaneous constraints and goals. Certainly, the analogy with the architectural domain suggest that, at a minimum, the assessment of exhibits is a multi-dimensional endeavor. At best, the metaphor may help us to critique, to analyze, and to gain insight into the whole issue of exhibit design.

Criticism

"The function of criticism should be to show *how* a work is. What it is, even that it is what it is, rather than to show what it means." (Susan Sontag)

"The end of criticism is the re-education of the perception of the work of art. The task of the critic is to lift the veils that keep the eyes from seeing." (Dewey)

Criticism, I think, is sorely lacking in the field of exhibit design and assessment. Perhaps this is because criticism is seen as an activity in which one sets out to expose the flaws and weaknesses of exhibit designs. This sole focus on the negative is not the kind of criticism that is referred to here. Rather, criticism as meant here is the discussion of and the illumination of the key qualities and characteristics of the exhibit that help others to understand more deeply the nature of that exhibit.

Artful criticism requires connoisseurship. That is, to criticize a piece well – to illuminate its essential qualities – requires an eye which is highly experienced. It also requires deep affection for one's field. Elliot Eisner, in his book *Educational Imagination* describes connoisseurship in this way:

"To be a connoisseur is to know how to look, to see, and to appreciate. Connoisseurship, generally defined, is the art of appreciation. It is essential to criticism because without the ability to perceive what is subtle and important, criticism is likely to be superficial or even empty."

Eisner goes on to describe the difference between connoisseurship and criticism:

"Connoisseurship is the art of appreciation, criticism is the art of disclosure. Connoisseurship is a private act; it consists in recognizing and appreciating the qualities of a particular piece of work, but it requires neither a public judgement nor a public description of those qualities.

Criticism is the art of disclosing the qualities of events or objects the connoisseur perceives. Criticism is the public side of connoisseurship. One can be a connoisseur without the skills of criticism, but one cannot be a critic without the skills of connoisseurship."

Critics provide us with "accurate, sharp, and loving descriptions." Critics also point out deeper underlying themes, ideas, or issues that permeate the entire work. Critics point out in a personal and subjective manner, the ways in which the work works for the audience (e.g. the ways in which the experience of the audience is heightened by interaction with that piece of work.)

In the movie *Amadeus*, Salieri, as an old man confined to an insane asylum, remembers with deep affection and explains to the listener the magical qualities of Mozart's music. He conveys appreciation and affection for the music, while at the same time pointing out to the listener the qualities that make the music great. It is a criticism that is highly educational.

On the opposite end of the spectrum from the agricultural paradigm, the metaphor of criticism might provide us with great freedom in helping us to identify, articulate, and debate the aesthetic and artistic characteristics of our exhibits. Not bound by the chains of "objectivity," the admittedly subjective art of criticism would allow us to draw upon the wisdom and expertise of those who have been working in the field—connoisseurs of exhibits and informal science education. Criticism, legitimized as an evaluation technique, might allow us to make formal and public what happens now in a frequent but informal and often private manner.

Criticism might help us identify qualities or characteristics of exhibits that will be successful or unsuccessful in an experiential way with the museum visitors. Good critics can generate hypotheses and important questions to be tested through close observation and working with public audiences. Good critics can help to identify "outcomes" never anticipated or thought of by the designer. Good critics may identify barriers to successful experience that can easily be eliminated. And finally, good critics should help raise the level of debate, discussion, and reflection on the art of designing educational exhibits.

Investigative Journalism

"The job of the investigative reporter is to examine all of the institutions of society and report how they work—not how they were designed to work, or how their leaders claim they work, but how they really work." (Williams, 1978)

Evaluators operating in museum settings probably do not need to operate with the same degree of suspicion as do investigative reporters operating in a political environment. Nevertheless, evaluators might profit by adopting the investigative journalist's goal of discovering information that is

hidden. The journalist is looking for information that is hidden with malicious intent. The evaluator, on the other hand, looks for realities that are hidden by surface descriptions, by blindness due to wishful thinking, or by a simple lack of awareness.

The approach and methods of the investigative journalist are perhaps most useful when the nature of the evaluation calls for a non-intrusive study of ongoing processes as they occur in their natural setting (e.g. visitors in a museum setting.) The approach may be useful when there are individuals or organizations not fully aware of how their programs or exhibits are functioning, and uneasy about discovering that the functioning of the exhibit is quite different from what it is presented to be or from what it is supposed to be.

The investigative journalist typically begins a project with a "fast study." This is the initial phase of the investigation where the reporter seeks to learn the territory. Through a brief immersion in the scene the reporter seeks to learn the laws and norms of the setting and to discover the perspective of the various parties involved. In the case of a museum evaluation, such a "fast study" might involve learning the perspectives of those who are planning and designing the exhibits as well as a quick observation of the actual use of those exhibits. The outcome of such a "fast study" would be a series of hypothesized discrepancies or "suspicions" that focus on those places where intention and reality diverge.

"Tracking" is one main technique used in the initial phase of the journalist's investigation. Starting with a hunch (hypothesis) the investigative journalist looks for evidence that confirms or denies his or her suspicions. The reporter uses all kinds of worldly knowledge in imagining where and how "tracks" would be left behind, if the hunch were true. In a museum setting the evaluator operating under the metaphor of the investigative journalist would form hunches about the reasons for the discrepancies between the designer's motives and the reality on the exhibit floor. The evaluator would then pursue through observations, interviews, and examination of recorded evidence reasons for those discrepancies.

Other journalistic approaches and techniques may also be useful for the museum evaluator. Modus operandi, the notion of boundaries, circling, shuffling, and triangulation may all have their analogues in the museum.

Also it is interesting to note that journalists do not strive for objectivity. They readily admit the impossibility of finding some sort of objective truth. Rather the criteria that journalists seek is "fairness." A fair story is one that presents all points of view, from both the advocates position and the detractors. Fairness might be an interesting concept to explore in the museum assessment domain.

Anthropology/Geography

There has been considerable work done already in museums under the anthropology/ethnography metaphor, combining techniques of observation and interviewing. The aim of this research approach is to understand in some sense

how the culture of the museum and the culture of the visitor are interacting. Meaning is always a function of context. Through observations and interviews the researcher can begin to infer what kinds of meanings visitors are creating out of their experiences in the museum. Such meanings tell us about the contexts in which visitors are actually operating.

In addition to the specific research techniques of the anthropologist, it is perhaps the heavy emphasis on the notion of culture that is the most important aspect of this metaphor. Not only is it the visitors that live in the recreational/educational culture of our society, but also the museum staff and exhibit designers are part of a professional/scientific culture that very much shapes their interests and behaviors. In some sense the mission of the museum is better described as "acculturation" than as "learning." If we look at the visitor's interaction with the exhibits, and with the whole museum environment, as an experience in entering a new culture, then perhaps anthropology can teach us some appropriate questions to ask and provide us with ways to pursue them.

Geographical research (particularly from the humanistic school) provides us with another closely related model of naturalistic inquiry. Geography focuses heavily on understanding human civilization in the context of the surrounding landscape. The fundamental aim of the geographer as a humanist is to develop understanding by revealing the richness and complexity of human experience. Unlike the experimental modes, the goal of this research is not to simplify or reduce multi-dimensional worlds to single dimensional worlds, but quite the contrary. The goal here is to even further enrich our models of the structural complexity and richness of the observed world. The goal in fact is to amplify and extend understanding beyond its initial (perhaps stereotypical) view. To gain such understandings it is important that the research expose the perspectives of the people involved and illuminate the context of their actions. This context is determined at least in part by the local environment (e.g. the museum, its staff, its resources, and its ambiance.)

Local landscapes not only determine, but they also reveal cultural meanings. Accordingly, careful documentation of museum environments (perhaps through photography or video tape) might tell us about the educational mission of museums in much the same way as do the pictures in National Geographic.

Finally, there may even be specific techniques of cartography and the analysis of spacial relationships used by geographers that may be useful for evaluators.

There are, of course, other metaphors that may be useful to us in thinking about alternative approaches to assessment in the museum environment. I will mention, very briefly, a few of these below.

Product Evaluation

Consumer Reports is probably the best known example of product evaluation. In exploring this metaphor for its

application to museum evaluation, one of the most interesting aspects of the product evaluation metaphor is its focus on consumer need and consumer satisfaction. Product evaluation is not particularly interested in the goals of the designer, but rather in the extent to which the product satisfies consumer needs along many dimensions. Every *Consumer Report* review contains charts that compare products across multiple criteria that themselves expand multiple dimensions. Such dimensions include: features and convenience, sensitivity, reliability, performance, cost, safety, and so on—depending on the nature of the product being evaluated. In some sense, *Consumer Reports* is simply a more systematic, measurable, and needs-based approach to connoisseurship and criticism.

Narrative – Storytelling

Some argue that human experience is essentially contained in the form of stories. A narrative that describes experience is close to experience itself. The function of narrative is to create unity, wholeness, and simplicity out of an otherwise chaotic flow of actions, events, and experiences. Narrative allows the almost direct transferral or recreation of experience. This is perhaps why the anecdote (although maligned in scientific circles) is such a powerful vehicle for conveying the nature of experience.

There are two devices used in storytelling that might be useful to the evaluator. One is the vignette, which is a short scenario giving us a powerful glimpse of reality. Vignettes, selectively chosen, allow the storyteller (or evaluator) to illustrate an important theme, feeling, or pervading issue. They are powerful in their specificity, but at the same time remain general.

Another storytelling device is dialogue in which the interaction between two people is captured and, like the vignette, can be both specific and general.

Capturing the stories of those who visit the museum may be a powerful tool in helping us to understand what they take away from their visit. Throughout their visit and as they leave, visitors construe their experience. They explain it to themselves and they remember their experience according to the story they've chosen about it. Evaluation may be able to use both the techniques and the structure of stories in order to better understand the internal experiences visitors are having.

Committee Hearings

In 1973 the Senate Watergate Committee gave the world a dramatic example of the use of a committee as an investigative and evaluative tool. Committees bring together, face to face, all those who have a stake in the evaluation: decision makers, evaluators, program personnel, clients, audiences, etc. Committees not only expose different points of view, but they allow interaction between those differing points of view.

The committee hearing metaphor, although not followed literally, may be a useful approach when a committee already

exists and is suited to overseeing the evaluation task. It is also a useful vehicle when a written report is likely to be ignored or have minimal effect. In essence a committee hearing combines the doing and the reporting of the evaluation.

A useful notion of the congressional committee structure is the position of counsel and staff. They are selected by and responsible to the chair person of the committee. In addition, they carry out the requests of the committee. In the evaluation analog the evaluator might play the role of counsel—collecting background data, carrying out small research projects, and performing initial interviews, in order to prepare the material to be covered in the “committee hearings.”

Marketing

The focus group is a marketing instrument that has received recognition as a widely applicable research tool. By assessing the satisfaction of consumers (e.g. visitors), the focus group allows for extensive probing of public interests and tastes. In a broader, but less deep fashion, marketing surveys can similarly be used to assess public tastes.

Cognitive Science

The cognitive psychologist, and more recently the cognitive scientist, have studied thinking by collecting and examining very detailed protocols. They record their “subjects” as they talk out loud describing their thinking as they solve problems. In reviewing the transcribed protocols, the researchers look for the fundamental structures of thought and the basic mechanisms used in problem solving processes. Similarly, such talk aloud protocol with detailed analysis may be useful in examining visitor's cognitive experience as they interact with exhibits.

Ethology

In her doctoral thesis, studying the behavior of families in science museums, Dr. Judy Diamond applied the methods

of an ethologist to the study of visitors in science museums. As she did when she studied the behavior of coyote families, Judy worked out a detailed protocol for codifying behavior so that the micro-interactions between family members were carefully documented. This approach allowed her to speak with great specificity about the kinds and numbers of interactions between family members under varying circumstances.

These various metaphors illustrate alternative assessment approaches on at least two levels. One, they illustrate alternative paradigms to the experimental (agricultural) paradigm. Some, like anthropology and geography, are based on a more naturalistic paradigm that relies on the techniques of field study and seeks to provide insight more than to prove absolute truth. Others are based on a more judgmental and subjective paradigm, such as criticism, which seeks to provide its audiences with language that helps describe aesthetic and personal experience. Some of the metaphors are based in an adversarial paradigm, such as congressional hearings and investigative journalism. They attempt through a process of cross-examination and triangulation to uncover that which is hidden and to provide a balanced picture, achieving “fairness” in evaluation.

Footnote

- 1 - I am indebted to the work of Nick Smith, and the Research on Evaluation Project, for many of the metaphors that are discussed in his paper. See, for example, Smith, N. (1981). *Metaphors for Evaluation*. Beverly Hills, CA: Sage Publications.
- 2 - The production of this paper was supported in part by a study to examine long-term relationships that visitors form with science museums undertaken by the Joyce Foundation, the Chicago Museum of Science and History, and eight other institutions—New York Hall of Science, Field Museum of Natural History, Franklin Institute Science Center, Chicago Academy of Sciences, and Discovery Place, Charlotte.

NEW PUBLICATIONS FROM THE CENTER FOR SOCIAL DESIGN

#90-10. *An Evaluation of Man in His Environment*, by Harris Shettel. [\$10.00]

Originally prepared for the Field Museum of Natural History, this report is part of a reprint series of difficult-to-obtain reports.

#90-20. *The Role of Simulated Immersion in Exhibition*, by Stephen Bitgood. [\$4.00]

This report discusses the experience of immersion and reports a research study on this topic.

#90-30. *Fish Stories that Hook Readers: Interpretive Graphics at the Monterey Bay Aquarium*, by Judy Rand. [\$2.50]

This is a reprint originally published in the 1985 AAZPA Annual Conference Proceedings.

For information on how to order these reports contact: Center for Social Design, P. O. Box 1111, Jacksonville, Alabama 36265