

**Note:** The following article is based on a presentation to the American Association of Museums Annual Meeting in Philadelphia, PA, 1995.

## Why Museums Don't Evaluate

Alan J. Friedman, Director  
New York Hall of Science

Only a minuscule percentage of exhibitions receive formal evaluation. This is true despite a long list of rational arguments in favor of evaluation, which readers of *Visitor Behavior* would have no difficulty assembling. For starters, communicating culture to the public has been part of the *raison d'être* for museums since the late eighteenth century, and this role remains a key element in the mission statements of every non-profit museum in the world. Museums cannot know if they are achieving that mission without rigorous evaluation of the effectiveness of their exhibits. In the presence of fundamental arguments like this one, why does the percentage of exhibitions which receive formal evaluation remain in the low single digits?

Science and technology centers are a good place to look for explanations of why museums don't evaluate. Of all the types of institutions in the museum realm, science-technology centers and science museums might have been expected to embrace visitor studies and evaluation the most warmly. After all, science itself consists largely of research and evaluation. Scientists are normally among the most critical of any educational, philosophical, or programmatic proposition which is advanced without a plan for rigorous evaluation. Science museum staff therefore ought to have a burning desire for visitor evaluation, a desire to know if the museum's exhibits and programs are doing what they were designed to do.

Yet most science centers and museums have just as little in the way of hard-nosed evaluation of their effectiveness as a medium for communicating with the public as do art and history museums. Here are several possible explanations:

- A fear of "no significant difference." Perhaps the most frustrating result of an evaluation study is that the program being tested produced no significant difference. In science this means either that the evaluation protocol is inadequate, or that the proposition being studied simply does not "work." Since refining testing methods and validating propositions are the daily work of science, the risk of "no significant difference" is an acceptable, common, intermediate step along the way to teasing out the truth (and gaining tenure). By contrast, exhibit evaluation is usually not included in the job descriptions of curators, exhibit directors, or designers. For them, "no significant difference" is often seen as proof of time and money wasted on evaluation.

- A fear of negative results. Most museum staff have seen some exhibits, perhaps even in their own institutions, that they suspect might actually produce negative results: visitors might learn or reinforce misconceptions through an exhibit, or visitors might find themselves less interested in a subject than they were before. While "no significant difference" might be frustrating and indicate a waste of time and money, a negative result would be directly bruising to egos, could endanger future funding, and might be damaging to careers.
- Distrust of educational evaluation in general. The possibility of "no significant difference" or a negative result are acceptable risks in science because scientists generally trust the evaluation processes they use to produce truthful and useful results fairly quickly. However, museum scientists generally have little knowledge or confidence in the tools that can be brought to bear for evaluating visitor behavior and learning. These hard scientists equate visitor studies primarily with marketing surveys and with the generally disappointing reputation of education research in the formal school system, where "no significant difference" is the most common result of comparing various educational treatments across different students and teachers. Without confidence that visitor evaluation can produce meaningful information, museum scientists don't want to run the risks evaluation entails. "No significant difference" or negative results in the realm of visitor studies probably demonstrate, most curators and exhibit designers prefer to believe, that evaluators don't know how to measure what their exhibits actually achieve.
- Lack of consequences for avoiding visitor evaluation. There is relatively little negative consequence for an exhibit or program which looks handsome but communicates little. By contrast, an exhibit which looks shabby, misses its opening date, or exceeds its budget can have immediate, harmful consequences for the museum's staff. So attention to appearance, timeliness, and completion of the promised square footage is usually a far higher priority than is attention to an exhibition's ability to communicate effectively.

Recently a large new museum discovered that visitors were staying longer than the planners had predicted. This seriously cut the museum's capacity to handle the intended numbers of visitors, and thus earned revenue was coming in at well below the amount budgeted. The consequences included deficits and severe reductions in the number of staff. It is hard to imagine that a failure to evaluate the effectiveness of the museum's exhibitions would have had any consequences as serious as those of the inaccurate forecast of earned revenue.

Evaluation is sometimes promised by a museum because a few funders look for an evaluation plan in proposals. The National Science Foundation (NSF) has been a leader in

demanding an evaluation component for every exhibition proposal. Nevertheless, outside of NSF's scrutiny as part of judging a museum's next proposal, there is rarely any follow-up by funders to see that evaluation was in fact done and what was learned. Funding officers, especially in the private foundation world, have just as few consequences as have museum staff if evaluation is simply avoided, and have just as much at risk as do the museum staff if the result of an evaluation is disappointing.

Establishing evaluation as a positive, even essential, component of the museum profession will require a significant change in the attitudes of most museum leaders and funders. Education and persuasion, aspects of political movements of all kinds, are in order.

#### **Educating Museum Scientists and Leadership**

Presentations of positive case studies by evaluators and staff from other museums, circulating publications about evaluation, holding short workshops on evaluation using a museum's own exhibits for practice – all these can demonstrate that powerful tools are available for visitor evaluation and that evaluation produces invaluable results. Testimonials from professionals who were persuaded to try evaluation, and now use it regularly, may be particularly helpful.

General arguments for evaluation, no matter how elegantly crafted, may be less effective than very specific and dramatic case studies. Consider the evaluation performed by Falk and Weiss (1993) on a computer-based exhibit produced by the New York Hall of Science on the subject of HIV and AIDS. The transmission of the virus is a matter of potentially deadly ignorance and misconception among the American public, particularly among sexually active teenagers. For example, there is a widespread misconception that any form of birth control, including the pill, provides protection against HIV infection. Evaluation quickly revealed that the HIV-AIDS exhibit unit did communicate some of the basic biology of the virus, but failed in its original form to make much of an impact on this widespread misconception. The resulting modification of the exhibit's program, however, made a major improvement in visitor understanding. Evaluation conducted after the changes were made showed that twice as many visitors could then answer correctly a question related to the relative efficacy of condoms versus other means of birth control in preventing the transmission of the deadly virus. The new version of the exhibit, improved by evaluation, has now been used by millions of visitors as a component of three large traveling exhibitions, three small traveling exhibitions, and one permanent exhibition.

Stories of successful evaluations which resulted in improved communication, money saved, mistakes avoided, and new projects funded can help create the positive desire for evaluation even in the face of the risks involved.

#### **Recognizing the Consequences of the Lack of Evaluation**

To some extent, the consequences of years of doing

without evaluation are just now beginning to be felt. The current political climate in the United States is cutting resources for any endeavor which does not have convincing evidence that it is effective in attaining high priority goals. Without far more visitor evaluation, the claims of museums to be important factors in producing better informed citizens, higher-achieving students, and a more competent workforce will not convince the critics who have better uses for the public and private funds which currently go to support museums.

Controversies at the Smithsonian Institution over the *Enola Gay* and *Science in American Life* exhibits demonstrate another consequence of the lack of evaluation as a routine component of museum practice. These exhibits were criticized for what the critics believed were the negative impacts they would have on visitors. The *Enola Gay* exhibit was attacked in part because, its critics claimed, it would degrade the public's patriotism and respect for World War II veterans (Wallace, 1995). *Science in American Life* was attacked because it was said to be too negative and would damage visitors' appreciation for the many real benefits of science and technology: Indeed, a visitor might come away convinced that science is a serious threat to human life (Park & Goodenough, 1996). One way to have resolved these criticisms would have been to evaluate the exhibits (or preferably prototypes) with actual visitors, to see if the exhibits did in fact produce the effects the critics predicted (Friedman, 1995).

Perhaps because visitor studies are not a routine part of exhibit development, evaluation had no opportunity to play an early role in resolving either of these conflicts. The original *Enola Gay* exhibit was cancelled before it was built. A much smaller and blander display took its place and the museum's director resigned. An exemplary summative evaluation of *Science in American Life* (Pekarik, Doering, Bickford, 1995), produced more than a year after the criticism had erupted, demonstrated that the exhibit in no way damaged visitors' appreciations of the positive benefits of science and technology. Nevertheless the critics (primarily scientists themselves) continued repeating precisely the same attacks, aware of, but apparently unmoved by, the rare appearance of empirical evidence in the realm of museum exhibitions (Shields, 1996).

Finally, "edutainment" profit-making enterprises are competing for much of the same audiences that museums serve, promising entertainment and cultural enrichment as well. Museum staff believe that they are more serious about education and communication, but are museums in fact making the case that a visit to a theme park is not a replacement for a visit to a museum? The lack of widespread evaluation results to demonstrate the efficacy of museums may make their case an increasingly hard sell. Convincingly differentiating museums from "edutainment" by means of evaluation may be one essential step in assuring the survival of museums as informal educators in the next century (Friedman, 1996).

### Achieving Better Discipline in the Exhibit Development Process

Once the concept of evaluation is clear and approved, the most common disincentive to performing evaluation is a lack of discipline in the exhibit development process. This lack makes evaluation difficult, expensive, and unproductive. Without a tightly controlled process, exhibit development always seems rushed, with no opportunity to pause for evaluation, and no opportunity to make use of what is learned from evaluation.

Exhibit development always seems to take longer and cost more than originally budgeted. When cost overruns occur, the temptation is to jettison evaluation in favor of getting more exhibit units completed. Michael Spock has described (see his article in this issue) meeting this challenge by holding a percentage of an exhibit budget in reserve for remedial evaluation after opening day.

The struggle to include evaluation among the daily tools of exhibit development has high stakes for the future of museums as well as the employment of museum evaluators. As described here, the issues and tactics should be treated not only as intellectual matters, but as political and financial ones as well.

### References

- Falk, J. H., & Weiss, M. (1993). Utilizing museums to promote public understanding of science: Early adolescent misconceptions about AIDS prevention. In *Visitor studies: Theory, research and practice, Vol 6*. Jacksonville, AL: Visitor Studies Association. pp. 98-105.
- Friedman, A. J. (1995). Exhibits and expectations. *Public Understanding of Science*, Vol. 4, 312.
- Friedman, A. J. (1996). Differentiating science-technology centers from other leisure-time enterprises, *ASTC Newsletter*, Vol. 24(1), 7-10.
- Park, R. L., & Goodenough, U. (1996). The unmaking of American science policy: The end of the scientific era? *Academe*, January-February 1996, p. 15.
- Pekarik, A. J., Doering, Z. H., & Bickford, A. (1995). *An assessment of the 'Science in American life' exhibition at the National Museum of American History*, Institutional Studies Office, Smithsonian Institution, Report 95-5 (November 1995).
- Shields, J. E. (1996). *Science in American life revisited*. *Chemical and Engineering News*, March 11, p. 40.
- Wallace, M. (1995). The battle of the ENOLA GAY, *Museum News*, July-August, pp. 40-45, 60-62.

**Note:** The following article is based on a presentation to the American Association of Museums Annual Meeting in Philadelphia, PA, 1995.

### Evaluation Climates and Conversations

Michael Spock  
Informal Learning Program  
The Chapin Hall Center  
at the University of Chicago

Not surprisingly, both fundamental and situational concerns surround the practice of museum exhibit and program evaluation. I would like to explore each of these levels of concern by first offering some observations on the situational politics of our recent exhibit evaluation work at the Field Museum, and then conclude by suggesting how semantics plays a more fundamental role in the way we relate to the evaluation process.

I left the Field Museum, a large collections and research-based natural history museum, in 1994 after eight and a half years of intense renewal work in which we renovated 140,000 square feet of exhibit space to the tune of more than \$25 million. And significant attention was given to the Museum's education programs as well.

A great deal, although not all, of this renewal was informed by evaluations of various sorts. How did we do these evaluations? How did they shape our work, our exhibits, our programs? What sort of climate for evaluation existed when we began? How did that change and why? What is happening now that things have slowed down and some of us have left the Field?

There had already been some exhibit and a lot of program evaluation when I arrived at the Field Museum of Natural History in 1986. The most recent big permanent exhibit, *Maritime Peoples of the Arctic and Northwest Coast*, had been extensively evaluated and Harris Shettel did a major summative evaluation of *Man and His Environment* back in 1975. The Education Department was systematically defining goals and objectives and evaluating most of their programs as I walked in the door.

So I was not coming into an evaluation desert. There may have been significant pockets of indifference to program and exhibit evaluation, but little detectable hostility. However, in a curatorially driven museum, I think it is fair to say that there was a much stronger commitment to the standards of content accuracy and object conservation than to the rigor of presentational efficacy in public programming.

The first politically relevant thing to understand is that I had a clear mandate from the board and senior management to take control of the direction and execution of public programming and make it work. Whether the implications of this mandate were fully understood, the board and senior