DOCUMENT RESUME

ED 182 159 SE 029 880

AUTHOR	Linn, Marcia C.
TITLE	Uses of Evaluation in Science and Technology Centers: Informed Decision Making. Advancing Education Through Science-Oriented Programs.
INSTITUTION	California Univ., Berkeley. Lawrence Hall of Science.
SPONS AGENCY	National Science Foundation, Washington, D.C.
REPORT NO	AESOP-PSc-17
PUB DATE	May 76
NOTE	20p.; For related documents, see SE 029 877-881 and SE 029 945-947; Contains occasional light and broken type.
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Demonstrations (Educational); Elementary Secondary Education; Enrichment Experience; *Evaluation Methods; *Exhibits; Museums; Science Activities; *Science Education; Science Programs; *Science Teaching Centers

ABSTRACT

Presented is a discussion on how an evaluator can function effectively in a Science and Technology Center, using an evaluation model on experience in evaluating exhibits at the Lawrence Hall of Science over a six-year period. The approach in this paper is described for four areas: (1) evaluation of currently available exhibits and programs; (2) selecting areas for new programs; (3) evaluation during development of an exhibit; and (4) evaluation of the learning that occurs. (Author/SA)

Uses of Evaluation in Science and Technology Centers: Informed Decision Making

by

Marcia C. Linn

AESOP* Lawrence Hall of Science
University of California
Berkeley, California 94720

May 1976

*AESOP (Advancing Education Through Science-Oriented Programs) is supported by a grant from the National Science Foundation

USES OF EVALUATION IN SCIENCE AND TECHNOLOGY CENTERS: INFORMED DECISION MAKING

Marcia C. Linn[1]
Lawrence Hall of Science
University of California
Berkeley, California 94720

Evaluation has gained a rather unenviable reputation. Just mention the word and defenses rise to the occasion. Some negativism undoubtedly stems from a feeling that evaluation will be destructive and discouraging. One often hears, "It's easier to criticize than create." Many program developers explain that they already have a large supply of criticism, "Just attend a staff meeting and see for yourself how critical everyone is!" Others note, "We don't have time (or money) for evaluation."

One model that has been applied to Science and Technology Center evaluation comes from work in curriculum evaluation. Traditionally attention has been focused on what students learn from a given program. Tyler (1951) as early as 1930 (see also Smith and Tyler, 1942) laid the foundation for curriculum evaluation as we know it today by insisting that program objectives be clearly defined before evaluation is carried out. Evaluation, generally as a comparison between two programs, became more common during the 40's and 50's. Cronbach (1963) pointed out the logical problems involved in comparing curricula which were intended to accomplish different goals. His major point was that tests to measure the objective of one program are unlikely to be covered as well by another program. More recently Scriven (1967, 1975) has emphasized the judgemental nature of all evaluation decisions and suggested that comparisons are essential for successful discussion making. Scriven has

---

[1]Paper presented at Association of Science and Technology Centers Workshop, "Measuring the Immeasurable," Baltimore, May, 1976. This work was supported, in part, by a National Science Foundation Grant to Advancing Education through Science Oriented Programs (AESOP). The author appreciates comments on an earlier draft of this paper from L. Eason, A. Friedman and M. Laetsch.

taken concepts from other fields and noted their usefulness in curriculum evaluation (e.g., goal-free evaluation as found in anthropology and cost effectiveness as used in industry). Stake (1967) has placed emphasis on individual differences in response to instruction. Recently Cronbach (1976) has pointed out that the many factors which influence program effectiveness include the society which uses the program; he argues that evaluation will never result in "laws," but rather must be constantly carried out to respond to the needs of a changing society. Many valuable and useful ideas are found in curriculum evaluation work. Criticism of curriculum evaluation approaches have centered on their exclusive emphasis on learning rather then, say, implementation or teacher preparation. Evaluation studies tend to be done in utopian sorts of environments where enthusiastic students and teachers follow every suggestion of the program developers. Additionally data collection methods have often been limited to paper and pencil tests.

In applying curriculum evaluation work to Science and Technology Center evaluation, we need to be careful to consider not only what is learned at these centers, but how to improve them, and how to bring visitors back. Unlike schools, Science Centers depend on voluntary visitors. Also, since the environment is informal, we need to find methods for collecting evidence which is relevant to the situation: Science Center visitors are unlikely to enjoy answering 20 multiple choice questions and will surely allow their annoyance to influence their performance.

I would like to discuss how an evaluator can function effectively in a Science and Technology Center. My model for evaluation has evolved from 6 years of experience evaluating projects at the Lawrence Hall of Science (LHS).

Evaluation is not a singular entity, hence it has many definitions. It encompasses many different techniques and can be applied in many different ways. The primary task of any evaluation is to select questions that need to be answered and adapt evaluation skills to answer these questions.

Lawrence Hall of Science Evaluation Model

An informed decision making model for evaluation is used at LHS. In science and technology centers evaluation is used to improve the programs, exhibits, and products so that they serve the target population better. Usually the target population is the general public; sometimes it is limited to school groups.

The users of evaluation information are the program and exhibit developers. What they need is information which will increase the likelihood that they will make effective decisions. First the evaluator helps the developers to discover all the possible decisions that could be made: create choices. Then the evaluator gathers information to facilitate making the optimum choice.

Certainly, one of the reasons that evaluation has not commonly been done is that it has not been useful. In some cases evaluators have answered questions irrelevant to the needs of the users. For instance, I could find out whether Lawrence Hall of Science visitors would prefer a Merry-go-round to the current science exhibits or whether visitors learn more at LHS than from a physics course. One reason that irrelevant questions have been addressed is that the informal learning environment itself is not well understood. Evaluators have looked for large gains in knowledge or changes in attitudes in voluntary visitors who come for a 2 hour visit. Clearly we need to choose our questions carefully.

The informed decision making model overcomes the problem of answering irrelevant questions by incorporating evaluation into exhibit development and gathering evaluation evidence for each major decision. When the evaluator is concentrating on choosing the best alternative, the positive characteristics of evaluation are evident. When the project developers and the evaluator agree on what to evaluate, it is unlikely that the evaluation will answer irrelevant questions.

Evaluators who assist in informed decision making function best as members of the development group. If challenged to prove the importance of evaluation, the evaluator can always gather evidence about some flaw in the program and "prove" that the program is unsuccessful. Clearly confrontation will not lead to communication. When this happens the project staff knows why they needed an evaluator--and why to avoid one in the future. The successful program or exhibit evaluator involves the whole development staff in evaluation. The project staff, in turn, relies on the evaluator to gather evidence for development decisions. Evaluation will generally save time and money when used for informed decision making.

External evaluation. It should be noted that in the type of program or exhibit evaluation discussed in this paper is internal evaluation. External evaluation can be very valuable and has been used effectively for early prototypes of programs and exhibits as well as for completed products but that is beyond the scope of this paper.

In this paper the evaluation for informed decision making approach will be illustrated in four areas: (1) Evaluation of exhibits and programs which are currently available; (2) Selecting areas for new programs; (3) Evaluation of programs and exhibits during development; and (4) Evaluation of what visitors learn.

Evaluation of Currently Available Exhibits

As noted above, the informal learning enviornment is not well understood. Some potentially meaningful variables have not even been considered. Many decisions in science and technology centers are made without much evidence. Exhibits might be spaced or clustered due to the availability of electrical plugs. Exhibits might be placed together because of similar exterior design, recency of aquisition or for any number of other reasons. Many times information which would increase the number of effective decisions could be easily gathered. Evaluation of currently available exhibits is useful for increasing our understanding of the informal learning environment.

We have been examining our current exhibits at LHS. Our findings illustrate some of the advantages of evaluating current programs. We know that the total average viewing time for traditional museum exhibits is 40 seconds so written material necessary for an exhibit should not require more than about 30 seconds to read. Our observations at Lawrence Hall of Science indicate that most users of participatory exhibits interact first and read when all else fails. When they do begin to read, they become frustrated if the text does not quickly give them the information they need. Alan Friedman reports that the format of the written material is very important; he found that visitors understood cartoon like directions for the astronomy exhibit much better than a printed statement using the same words.

While the Fossil and Minerals Exhibits at LHS are viewed for an average of 40 seconds, the puzzle tables are used for close to 5 minutes on the average and the computer terminals for an average of 16 minutes. From this evidence we can conclude that the amount of time spent at an exhibit is directly related to its participatory possibilities.

Recently Rita Peterson surveyed what visitors report liking about exhibits and what they feel they have learned. Thirty-four students aged 11 to 18 were asked to write responses to exhibits that interested them. Perhaps the most striking outcome of this survey is the wide range of comments for the same exhibits. Responses to the computer games, for instance, range from "Great," "Good Game" to "No point," "Not so fun," and "Crazy." Respondents indicated what they thought they learned from each exhibit about half the time. Comments ranged from "Nothing but still fun," "I didn't understand, but fun to play with" to "How a sundial works," or "Trial and error method doesn't work as well as systematic problem solving." The most compelling conclusion that we can draw is that visitors are individuals and that they come to LHS with a variety of interests and expectations.

A finer grained analysis of Rita Peterson's survey also tells us that participatory exhibits generate the most comments, but that the first exhibit the visitor encounters frequently elects comments (Minerals in this survey). We found that many 17 year olds reported that exhibits were too complicated or directions were confusing. A few visitors gave useful suggestions for improvements (e.g., would help if the holes in the wood puzzles were bigger). The most positive comments were associated with games of skill such as puzzles, reaction time, catching the pinball, or computer interactions. Machines which demonstrated complicated principles were viewed most negatively.

A common goal of visitor surveys is to find out which exhibits are most "popular." This is far from a simple question. It can be measured in terms of number of users, hours of use, number of return visits, liklihood of being remembered or many other ways. Our observations in the exhibit halls at LHS indicate that popularity should also be evaluated in terms of who the users are. In the past we have also used number of breakdowns as an indication of exhibit popularity.

Another goal of visitor surveys is planning for future exhibits. Some clear guidelines are available for this area. We know that the necessary written material for an exhibit should be simply expressed. We know that participatory exhibits are viewed longer than static or button pushing exhibits. We know that games which give the user a challenge and have a reasonable probability of solution are preferred to other exhibits by many visitors. We also know that our visitor audience is extremely diverse. Different exhibits will appeal to different sectors of this audience. As usual, variety is the spice of life.

Questions of interest for evaluation of exsisting exhibits include, who is the user audience?, What do they do?, Why do they come?, Why do they return?. The major sources of information to answer these questions are visitor surveys, staff feedback, and observation in the exhibit area.

Information can be used (1) to make minor changes to existing exhibits (e.g., provide stools so small visitors can see an exhibit, (2) to decide which exhibits to withdraw when there is crowding (the unpopular ones), (3) to develop new exhibits matched to the visitors that are now coming, (4) to design new exhibits which will make the center more appealing to a particular group, and (5) to plan an effective publicity campaign.

Selecting Areas for New Programs

One frequently hears that programs should be developed to meet the needs of the target population. This is a question which rightfully falls to the evaluator and certainly responds to evaluation techniques. The decision under consideration is "What new programs should be developed?" The techniques are first to create choices then to gather evidence to choose between them.

The task of creating choices cannot be underestimated. If the evaluator simply asks visitors what they want, visitors usually do not invent creative new exhibits on the spot so suggestions fall into the "more of this" category. Visitors <u>can</u> tell evaluators which of several new ideas they like best, why they come to a science center, and what existing conditions encourage them to return. The evaluator can also determine which segments of the visitor population are most interested in a given program and suggest how the center could attract more visitors of this sort. Thus the evaluator can help plan successful new exhibits by determing which of several ideas would work best.

Data comes predominantly from interviews with users when one is selecting new areas for program development. Evidence is used to plan new programs. An example is the Outdoor Biology Instructional Strategies Program developed at Lawrence Hall. A telephone survey of every type of community group was conducted to determine how these groups would incorporate outdoor activities. We found that most groups preferred activities which could be used alone or grouped at the discretion of the leader. This evidence was influential in the development of our final activity format: We developed individual activity folios instead of units composed of several activities.

Evaluation for Exhibit Development

Once a new exhibit or program is in the planning stage, an evaluator can help to create choices and gather evidence about the potential value of each choice. During the development of our new astronomy exhibit, for instance, each version of the exhibit was developed in preliminary form. The exhibit developer and other members of the staff asked visitors to try out the exhibit. Observations and interviews were used to assess user relations.

Figures 1 and 2 show preliminary and final telescopes for the astronomy exhibit. As a result of many comparisons of different telescopes and visitor interviews, Alan Friedman found that visitors expected the spyglass type of telescope rather than the angled telescope used by astronomers, visitors could locate stars better with the spyglass, and that visitors did not mind the slight discomfort involved in looking up rather than down. The different telescope heights were chosen to accomodate various sizes of visitors. More detailed information for the visitor who is interested was placed behind the interaction structure. This avoids having visitors standing in front of a telescope reading about astronomy.

In exhibit development the major question is how can the exhibit communicate better. Frequently the evaluator can help create choices by isolating aspects of the current plan which cause confusion. For example, in revising the optics exhibit we found that users had difficulty focusing lenses and changed one exhibit to emphasize focusing (Eason and Linn, in press).

Sources of evidence for decisions are usually observations and interviews with exhibit users. The evidence is used to make the exhibit communicate more clearly.

Evaluation of What Visitors Learn

When evaluation has been used in Science and Technology Centers it has generally been to determine what is learned. It is clear that visitors do not come only to learn. In fact, responses to Rita Peterson's survey indicate that visitors frequently don't respond to questions about what they learned yet clearly report enjoyment, interest, and a desire to return. Also, quantum jumps in knowledge do not take place during a 3 minute exposure to an exhibit or even a 2 hour visit to a Science Center. Visitors may become interested in a particular question and decide to pursue it further as the result of a visit to a Science Center or they may just be pleased to have challenged their abilities by doing a puzzle. We do know that learning is much more likely to take place when the learner is actively involved in learning (Thier and Linn, 1976). Our surveys at LHS indicate that this may well be related to increased time spent with materials which permit interaction.

In evaluating what is learned, the traditional models of curiculum evaluation, as described above, are often invoked. Tyler's (1951) concept of determining the goals of a learning experience, Popham's (1971) emphasis on criterion referenced testing, or Scriven's (1975) goal-free evaluation may also be used. It has recently been said that choice of an evaluation theory is dependent on its charisma coefficient. The charisma of the current flag bearer for the particular point of view determines which viewpoint will be followed. The availability of these and other theoretical viewpoints indicates the recent interest in evaluation. This interest has greatly expanded the possibilities for gathering evidence in an evaluation framework. At the implementation level the task of the local evaluator is to select appropriate techniques from all those available and to apply them effectively. Again, choices must be based on the decisions that need to be made. Tools appropriate for answering them must be found.

In our work at LHS we have developed some guidelines for evaluating learning in an informal environment. Clearly learning is only a part of any visit to a science center. Perhaps not so clearly it is not the only goal for such a center to achieve. Additionally, evaluation of what is learned is not as useful for improving the science center as evaluating what will make an exhibit communicate better or what will bring a visitor back. There are, of course, many excellent reasons for finding out how best to impart knowledge to visitors. One goal may be to inform the public about, say, the sources of energy. Another might be to alert visitors to dangers in their environment. A science center might familiarize visitors with important scientific machinery such a computers or mass spectrometers.

In keeping with the concept of informed decision making, evaluation of what is learned in a science center can effectively be linked to a question of importance to exhibit developers. Laurie Eason and I, for instance, compared machines with activity booths when we looked at the learning potential in optics exhibits (Eason and Linn, in press). In evaluating the astronomy exhibit, Alan Friedman is interested in whether providing class sessions before or after exposure to the exhibits is most effective.

Once the question is chosen, how can evidence be gathered to answer it? The first step for the staff evaluator when finding out what might be learned from a particular exhibit is to list the possibilities. Not just the goals stated by the exhibit developers, but other goals which could be accomplished as well. Usually the best way to determine these goals is to observe the exhibit in use.

Once goals are established, ways to assess them must be determined. Two major approaches are available: observation and visitor responses to written or verbal questions.

Observation

By observing exhibit users one can determine how long they spend with the materials, whether they complete the experience, what order they carry out the activities in, whether they leave and return, whether they talk to other visitors, etc. This information does not directly indicate learning but it characterizes the conditions of learning. It can indicate that certain learning could not take place if, for instance, no visitors complete the experience.

Direct Assessment of Users

The most common approach to assessing learning is to ask the users to answer questions. This involves (1) designing evaluation measures and (2) data collection and experimental design. Each of these topics will be discussed.

Designing Evaluation Measures. Use evaluation techniques that measure changes likely to result from the program. It is useless to measure reading achievement if the program doesn't teach any skill related to reading. Frequently it is necessary to design evaluation activities for the program, rather than depend on available measures.

Evaluation activities should be in the same mode that is used for learning. If the users learn from doing activities then they should be evaluated by doing activities. An example is the Museum exhibit evaluation study where we developed evaluation interviews with questions which matched two different types of exhibits (Eason and Linn, in press). We found that the questions most closely related in format to the format of the subject's experience were answered more frequently by students who had learned the material in that format.

Another application of this concept is that if students in the program do not write things down, then paper and pencil evaluation measures involve transferring learned information to the written mode and are less likely to be successful than tasks matched to the learning mode (Falk, Malone, and Linn, 1975).

It should be noted, however, that evaluation measures which involve jargon or conventions from the exhibit itself are likely to be easier for people familiar with the program. This sort of measure used with people who have observed the exhibit and those who have not will not convince external observers. Nevertheless, such an approach is useful to gather information for exhibit developers who want to know whether the users understand the jargon and conventions in the exhibit.

Developing evaluation activities can be as complicated as designing the exhibit. Evaluators need to follow up ideas for evaluation activities no matter how impossible they seem, and adapt them to the realities of the situation. If an idea for evaluation involves building a whole exhibit or asking 100 questions, the evaluator can think about it carefully, decide what the essentials are and then adapt the idea to reality. One example is an interview for the museum evaluation which was time consuming. Instead of omitting it, we interviewed only a small proportion of the students (Eason and Linn, 1976). In another case we gave only part of an interview to each child. In others studies we have built complicated prototypes of apparatus and then discovered simple substitutes which did the same thing.

<u>Data Collection and Experimental Design</u>. Data collection in evaluation studies is likely to pose problems. It is seldom possible or desirable to

set up a totally controlled experiment such as was originally done in agricultural research. It might be reasonable to grown half your beans in the dark, but museum visitors cannot be asked to spend their visit in an empty room. Many techniques have been devised to get around this problem. For example, we have interviewed half the visitors before they viewed the exhibit and half afterward. We have interviewed visitors before they view the exhibit on one weekend and after they view on another. Rather than comparing visitors who have had the experience to those who have not, we recommend comparing two ways of accomplishing the goal. Otherwise, those who have not had the experience are unfamiliar with the format of the questions and may require lengthy explanations of jargon or may differ in some other rather non-essential way from those who have had the experience. As noted above, even when two exhibit formats are compared, one may find that question format interacts with exhibit format.

We are currently exploring ways to avoid group comparisons in evaluation and look instead for evidence that we can interpret on its own. One way is to look for events which have previously not occurred. For instance, a new puzzle at the game table which never solved when first tried, but is frequently solved after the user has solved other puzzles. As mentioned above, we are also using observational approaches.

Our basic plan is to gather information from as many different sources as possible. We are combining observational data, situational data, interviews, and partially controlled comparisons. We anticipate that by finding out in two different ways that the same thing is true, confidence in the conclusion will be increased. Additionally, confidence in a partially controlled comparison is enhanced when other information supports the conclusion. For example, comparisons of intact groups using a new program do not control for group membership. If evidence from leaders, outside observers, and several interviewers all point in the same direction, then the conclusions are more convincing. Levine (1975) has called this the adversary model. Inevitably, when this approach is taken some information is inconsistent with other information. This is also of interest. Sometimes attempts to reconcile such inconsistencies result in great increases in understanding. Sometimes one is left with reporting the inconsistency and waiting for more information. For instance, the Peterson survey at LHS indicated that some children loved an exhibit while others found it boring. More detailed analysis may indicate that length of exposure to the exhibit, age, previous science courses or some other variable explain this feeling.

Thus, when evaluating learning in the science and technology center, it is useful to focus the investigation on an important question and then gather evidence to answer the question. An approach which generates evidence from many different sources is especially useful.

Summary

This paper has briefly suggested how evaluation done by on-site staff can serve science and technology centers. An informed decision making model was proposed. Use of the model in assessing currently available exhibits, center needs, development of new exhibits, and learning in the center was discussed.

Major questions which deserve attention in center evaluation are: (1) Characterization of the informal learning environment; (2) Assessing areas where the informal learning environment has not yet had an impact; and (3)

Devising methods for collecting evidence which does not depend upon comparisons between those who have had the experience and those who have not had the experience.

References

Falk, J., Malone, L. and Linn, M. C. Outdoor Biology Instructional Strategies (OBIS): Development and Evaluation. American Biology Teacher, 1975, 37, 162-165.

Levine, M. Scientific Method and the Adversary Model, American Psychologist, 1974, 29, 661-677.

Popham, W. J. (Ed.) Criterion-Referenced Measurement, Englewood Cliffs, N.J.: Educational Technology Publishers, 1971.

Scriven, M. "The Methodology of Evaluation." In Perspectives of Curriculum Evaluation. AERA Monograph Series on Curriculum Evaluation, No. 1 Chicago: Rand McNally, 1967, pp. 39-82.

Scriven, M. "Evaluation Perspectives and Procedures" in Evaluation in Education, J. Popham (Ed.), McCutchan, Berkeley, 1975.

Smith, E. R. and Tyler, R. W. Appraising and Recording Student Progress, New York: Harper, 1942.

Stake, R. E. "The Countenance of Educational Evaluation," Teachers College Record, 1967, 68, pp. 523-540.

Thier, H. D. and Linn, M. C. The Value of Interactive Learning Experiences in a Museum. Paper presented at Education and Science Centers Workshop, Association of Science and Technology Centers, Boston, October, 1975.

Tyler, R. W. "The Function of Measurement in Improving Instruction." In E. F. Lindquist (Ed.), <u>Educational Measurement</u>, Washington, D. C.: American Council on Education, 1951, pp. 47-67.