Using Evaluation to Develop Interactive Multimedia Exhibit Elements

Britt Raphling
Adler Planetarium and Astronomy Museum*
Chicago, Illinois

Introduction

Interactive multimedia exhibit elements seem to promise an unlimited palette of technical, aesthetic and interpretive features. For example, the user interface can be a keyboard with mouse or trackball, laser- or voice-activated, or a touch screen with an impressive array of button icons. A single interactive multimedia element can contain an apparently infinite amount of stored information, accessed by complex linkages. Screen format can be comprised of graphics, text, video, animation, sound and often, all of the above. Finally, there are many memory and speed options which can seriously affect the end product, depending on the complexity of graphics and animation.

But this incredible palette of choices can create as many problems as it solves. It is difficult to know how to tailor the extraordinary complexities of interactive multimedia elements to a specific message and audience. The array of features increases the odds that these exhibit elements will be technically confusing and educationally ineffective. Even more perhaps than traditional exhibit elements, interactive multimedia can truly overwhelm casual museum visitors.

Evaluation is especially useful to help focus interactive multimedia options so that the finished element will meet both exhibit content objectives and visitor experience objectives. This paper outlines an "ideal" evaluation process tailored specifically to the challenges of developing interactive multimedia elements in museums. It requires exhibit developers, programmers, visitors and the evaluator to participate in a series of five steps designed to ensure the most effective possible outcome. Although the evaluation methodologies described here are not new in and of themselves, what makes them especially suited to developing interactive multimedia is that they are used together in a specific sequence which tests, makes recommendations, revises and re-tests.

Using several evaluation techniques to develop a single interactive multimedia element may seem like costly overkill. Yet a more cursory evaluation is simply inadequate for these complex exhibit elements which

deserve more intensive scrutiny. It is surprising to learn, however, that even an extensive evaluation is not overly expensive, if it is integrated into the project from the planning stage (see note 1). The best way to guarantee that the time and money allocated to developing interactive multimedia are well spent is to use evaluation as a development tool *before* installation as well as afterwards. This helps ensure that more money and time won't be needed when it's too late to "fix" what visitors ignore or don't understand. When evaluation is integrated with the design process, it becomes an important development tool, and not just a check at the end.

Five Steps to Evaluating Multimedia

The "ideal" process has five parts, four of which occur before installation: (1) deciding whether to use interactive multimedia; (2) conducting a test with a paper mockup of the program; (3) doing a critical review of the first programmed version; (4) doing formative tests of the program with cued visitors; and (5) doing summative or remedial evaluation (after installation). Figure 2.1 is a flow chart that illustrates this five-part sequence.

Deciding Whether to Use Interactive Multimedia

Interactive multimedia elements do have qualities that set them apart from traditional exhibit formats, but for the purposes of concept development, they should be treated like any other exhibit element. Often multimedia are included in an exhibit because they are flashy and appealing, not because they are appropriate to exhibit goals. Team members can guard against this temptation by being explicit about why they need to use interactive multimedia and why other formats won't work as well. They need to set educational objectives and communication goals for users, and decide on a target audience. Also, there ought to be a conceptual link between the multimedia format and the content it is attempting to convey. For example, at the Museum of Science and Industry in Chicago, several multimedia interactives were created as part of an exhibit on high-tech scientific imaging. Because these imaging technologies are computer-based in "real life," it was conceptually appropriate for the museum to use multimedia simulations in the exhibit. A less appropriate use of interactive multimedia in an exhibit would be to have an entire museum collection contained in the program. This type of reference database would be more suited to a research setting, such as a library, collections room, or Internet web site, rather than the display environment of an exhibit.

Conducting a Test With a Paper Mockup

Once a concept has been developed for the interactive multimedia element, team members should create a paper mockup and test it with visitors. Dr. Barbara Flagg (see note 2), who has written quite a bit about this methodology, describes it as a "trial technique in which testers acted as the computer, presenting . . . paper screens while providing oral narration of what museum visitors would see or hear from the exhibit. When choices [are] available, respondents [discuss] what they [expect] to see/hear/or do with each choice. . . . An observer records the pathways visitors take and also all verbal and non-verbal responses to the materials." This paper mockup is not merely a design story board, but a paper version of the proposed program, with individual pages representing the individual screens. Visitors use the paper version as if it were the real thing, and evaluators assess their reactions to it.

This paper model can be simple or complex. Screens can be laser printed, hand-drawn or xeroxed. Videotapes of animation clips or "quick-time" movies and audio tapes of sound effects can be played for visitors at the appropriate time. One person acts as the computer, manipulating the "screens" and other features, while the other person makes observations and conducts the interview. Far from being distracted by this makeshift version of a high-tech device, visitors, especially children, will have fun playing along with the mockup computer. Because the program is literally still on paper, visitors can see that development is in process and that their opinions really count.

The inexpensive paper mockup gives a clear idea of where visitors "go" in the program, how they react to graphic aspects of its interface, and what they get out of it. Developers can then use this information to provide programmers with more specific guidelines as they begin to work on transferring the concept to an electronic format. It can also be useful to give the programmer a copy of the paper model as a starting point.

Doing a Critical Review of the First Electronic Version

Evaluator and developer should conduct a critical review before putting the program out on the floor for visitors to try. For the purposes of this paper, a critical review is a critique by an audience advocate, the evaluator, who is familiar with how museum visitors respond to interactive multimedia in exhibits. Critical review provides an opportunity to determine specific questions that will be evaluated. Because the actual programmed version will inevitably be different from how it was imagined by developers and how it seemed on paper, a critical review is an excellent first opportunity to assess the overall effect of how the concept has been translated into a programmed version. It is also a good time to do one more edit of text, graphics, or design. Comments from the critical review should go to the programmer for revisions.

Conducting Formative Tests of the Program With Cued Visitors (see note 3)

Once graphics, content and navigation mechanisms are in their "final" form, the program can go out on the floor to be tested with visitors. For this part of the evaluation, a small sample (10-20 visitors) from the target audience is asked to try the prototype. It is still a prototype, however, and doesn't necessarily need to be in a kiosk or have contextual graphics. Ideally, the program's interfaces work properly and graphics at least approximate the element's final look and feel. It should be complete enough so that it can be tested reliably, but still flexible enough to allow for changes in response to visitor feedback. The computer can be wheeled out on a cart, or visitors can be brought into a quiet space to use it.

Although the sample for this stage of the evaluations may be small, it is important that data be collected systematically, so that inferences made from the study will be valid. That is, the data should be gathered as consistently as possible with a specific interview protocol, in order to better avoid bias. The interviewer should present exactly the same demeanor to every visitor—right down to asking the questions in the same way and with the same tone of voice. This does not mean that the evaluator should be an automaton, merely that he or she should try to keep any sort of personal or leading responses from entering his or her interaction with visitors.

For formative testing of exhibit prototypes (both multimedia and traditional) it is helpful to use "cued" visitors, who agree to try out the interactive and give their comments. Testing cued visitors allows evaluators to assess the potential of an exhibit element—that is, if visitors chose to use the element and were motivated to figure it out, what could they get out of it? Cued visitors are often more motivated than uncued visitors, because they tend to spend more time, they put in more effort and often are more thoughtful respondents. However, they are not transformed into wonder-people who do everything, understand everything and like everything by virtue of their being cued. Cueing does not radically alter normal visitor responses, but simply amplifies them, whether positive or negative. When analyzing data generated from a cued sample, evaluators should set their standards for success slightly higher than for uncued visitor data, to account for that amplified response. For example, rather than considering a simple majority to be an indicator of success, an evaluator can look for 75% or 80%.

The formative study should (a) note how much time visitors spend with the program; (b) track which screens visitors use and in what order; and (c) include questions to assess how users understand and experience the program. Figure 2.2 is an example of a formative interview data sheet. What questions are asked during the interview will depend on what the exhibit goals are, and may also reflect other issues that arose during the critical review.

When the formative evaluation phase is complete, the evaluator should discuss results and recommendations with the team. This discussion should

include the programmer, because the more he or she knows about how formative evaluation works to guide exhibit development, the easier it will be for him or her to support program revisions that are based on and responsive to evaluation data. In fact, it is ideal if the programmer can observe some of the evaluation as it is conducted with visitors, so that he or she will understand concretely why evaluators make the recommendations they do.

If time allows, it is wise to do a second round of formative testing to assess the revised program with a new sample of cued visitors. This cycle (testing, revising and re-testing) can be repeated until team members feel comfortable with the program, or until time runs out. This repeatable, iterative loop is probably the most important aspect of the "ideal" scenario described here. Even if it is impractical to do any of the other evaluation stages, two rounds of formative interviews can radically reshape any interactive multimedia exhibit element for the better.

Conducting Summative or Remedial Evaluation

After the exhibit opens, it is time for a summative assessment, part five of this evaluation process. Such an evaluation can either be a remedial study (see note 4) which recommends changes when there are funds, time and interest left to make them; or a summative study which provides a final analysis of how the finished element performs with casual visitors in its actual environment. This summative or remedial stage draws on two different studies: tracking and timing, and an exit interview.

Tracking and timing allow evaluators to observe how randomly selected, casual visitors use the program. The data can reveal who users are, how much time they spend, what screens or parts of the program they look at or use, what percentage of the total program they use and whether they use it alone or in social groups. The tracking and timing data sheet consists of a flow chart of the program, similar to the one that is part of Figure 2.2, and places to note time spent and the subject's demographic characteristics.

The exit interview, conducted with a separate sample of randomly selected, casual users, gathers data on these visitors' attitudes, opinions and beliefs about the exhibit that is being tested. The evaluator observes a user from a slight distance and keeps track of how long he or she spends with the interactive multimedia element. When the subject is finished, the evaluator approaches with questions about content or anything else that can help determine what he or she got out of the experience. Timing how long subjects use the program is critical, because when visitors use an interactive element for less than 15 seconds, they tend not to remember the experience, which makes interviewing problematic. Figure 2.3 is an example of an exit interview data sheet.

Conclusion

With the first four evaluation steps completed (determining a rationale for interactive multimedia, paper test, critical review, and formative test), developers should feel that they have addressed most major problems and that the interactive multimedia element is as ready as possible for exhibit opening. The fifth step, summative testing, can give developers concrete information on actual performance in the intended environment. They can contrast this data with data from the prototype: for example, certain things may never have been corrected, or new problems may have arisen due to differences between the final format and the prototype, or due to the surrounding exhibit environment. Developers can also draw on information about both cued and casual visitor behavior to help them form a more realistic picture of how interactive multimedia elements are received in museum exhibits.

Developing interactive multimedia exhibit elements with prototyping evaluation as described in this paper is most appropriate for troubleshooting and cannot predict behavior or guarantee a perfect final product. However, because so many things can go wrong with complex multimedia interactives, this type of evaluation is critical for developing an effective exhibit element that can both inform and give pleasure to its intended audience. Many museum professionals shy away from evaluation because it is time consuming and may require them to revise their initial exhibit ideas; yet experience shows that we cannot trust our professional judgment alone. We need visitor responses at all stages of the development process so we can see clearly what works and what doesn't, and decide how best to fix it.

**Britt Raphling is currently Evaluation Manager at the Museum of Science and Industry in Chicago, Illinois.*

Notes

1. At the Museum of Science and Industry in Chicago, the budget for a multimedia project which developed and tested six programs was slightly over $1,000,000. A significant evaluation phase that included many of the steps outlined in this paper came to slightly over $5,000, or about .5% of the total budget. See Serrell and Associates' 1994 unpublished study for the Museum of Science and Industry, Chicago, "Computer Interactives Evaluation Final Report for the Mystery Lab in 'Imaging: The Tools of Science' " for a full description of the project.
2. Flagg, B. N. (1990). *Formative Evaluation of High Level Videodisc Design for "Beyond Earth. . .A Space Adventure": Report for Interactive Video Science Consortium.* Research Report No. 90-001. (Bellport, NY).
3. Technically, all methods of evaluating prototypes for exhibit development fall under the category "formative." See Griggs, S. (1981) "Formative evaluation of exhibits at the British Museum," *Curator*, 24/3, pp. 189-202; McNamara, P. (1987) "Visitor Participation in Formative Exhibit Evaluation," *Journal of Museum Education*, 12/1, Winter, pp. 9-11; Screven, C. (1976) "Exhibit Evaluation, A Goal-Referenced Approach," *Curator* 19/4, pp. 281-282; and Taylor, S., ed. (1991) *Try It! Improving Exhibits through Formative Evaluation* (Washington, D.C.: Association of Science-Technology Centers) for more on the application of formative evaluation to exhibit development.
4. For more on the differences between summative and remedial evaluation, see McLean, K. (1993) *Planning for People in Museum Exhibitions* (Washington, D.C.: Association of Science-Technology Centers), pp. 75-76; Bitgood, S. and Shettel, H. (1994), "The Classification of Exhibit Evaluation: A Rationale for Remedial Evaluation," *Visitor Behavior*, 8/4, Winter, pp. 4-8; Bitgood, S. (1995) "Introduction and Issues in Summative Evaluation," *Visitor Behavior*, 10/3, Fall, p. 4.

Figure 2.1: Flow chart depicting the five steps involved in using evaluation to develop interactive multimedia exhibit elements.

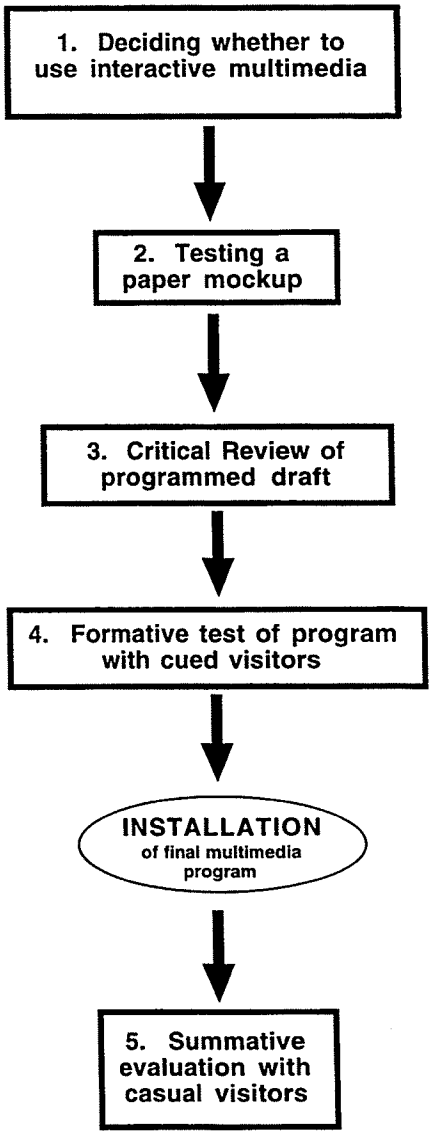

Figure 2.2: Example of a formative interview data sheet that was used with cued visitors to develop an interactive multimedia program.

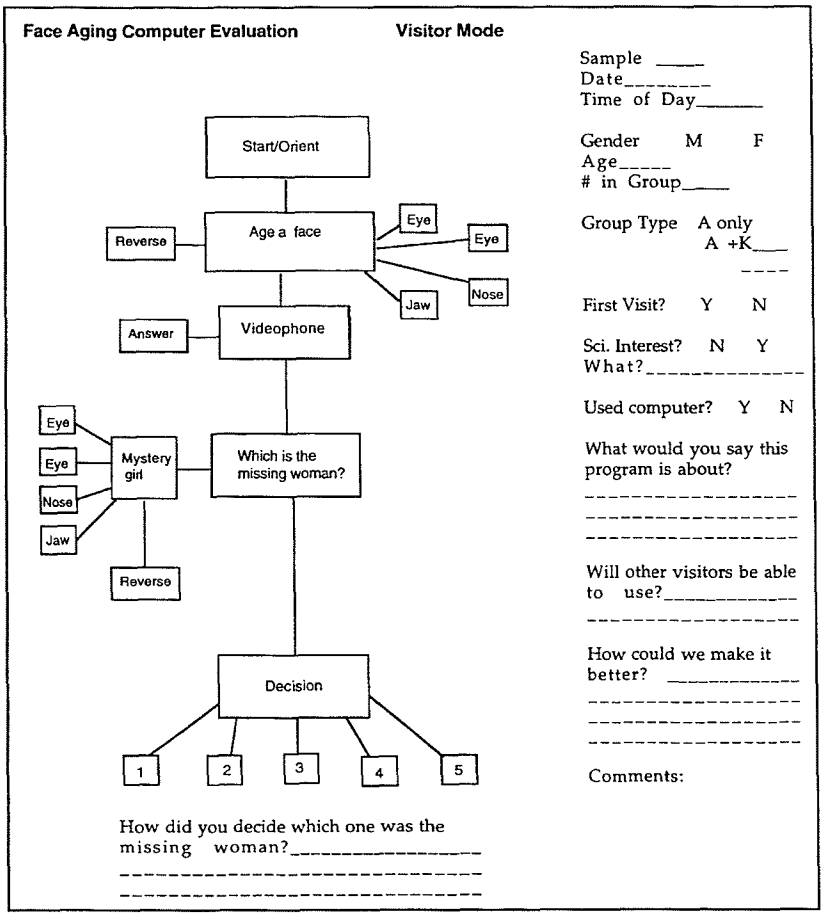

Figure 2.3: Example of an exit interview used for a summative evaluation of an interactive multimedia videodisc program.

Mule Deer Video Study--Summative Exit Interview

Date_____ Time of Day_____ Sheet #_____

Sex: Male Female Age: -20s 40s # Group size: Group type:
20s 50s 1 3 Adults only
30s 60s+ 2 4 5+ Adults & kids

Total Time spent _____ :00 - 1:00 1:01-2:00 2:01-3:00 3+ mins.

********** **********

(Intercept visitor a few cases away)
"Hello, excuse me, we're doing a special evaluation today in this exhibit, and I'd like to ask you a few quick questions."

Is this your first visit to [X] Museum?
Yes_____ No_____

Do you have any special interest in [exhibit subject]? No_____
Yes (describe)_____

Did you use that video over there? [Point to video]
Yes_____ No_____

Do you recall what it was about?

<probe to exhaustion>

Thank you very much for your time. Initials_____