# From "Guerrilla" Methods to Structured Evaluations: Examples of Formative Web Design from the Exploratorium's Evidence and Mind Projects

Sherry Hsi, Joyce Ma, Adrian Van Allen, Kristin Sikes, Melissa Alexander,
Exploratorium, USA

*http://www.exploratorium.edu*

## Abstract

This paper presents formative Web testing and evaluation methods created at the Exploratorium, ranging from floor testing with museum visitors using low-cost "guerilla" methods to structured evaluations that engage museum visitors and on-line remote audiences in the design process. This paper illustrates these methods using examples from two U.S. National Science Foundation-funded informal science education projects. The first project, "Evidence: How Do We Know What We Know?," addresses how to improve the public's understanding of a key element of the scientific process: how scientists attempt to construct a functional understanding of the world by gathering evidence, and how they make discoveries based on evidence. The second project, "Mind: Attention, Emotion, and Judgment – How do minds figure out what to do?," addresses developing a companion Web site to a new exhibition and educational programs that support museum visitors' exploration and experimentation with their own minds. In both of these projects, close collaboration between Web developers and evaluators enabled the collection of useful feedback. Different configurations of interactive media, the Internet, the museum floor, and visitors were used in concert to disentangle open-design questions and to generate constructive feedback and Web design revisions. This partnership approach and these evaluation methods offer developers new alternatives to standard usability testing approaches for improving Web site designs in the early development stages of complex projects.

Keywords: Web site evaluation, evaluation methods, scientific evidence, mind, user testing, science education

## Introduction

The Exploratorium's philosophy embraces a culture of supporting curiosity, play, and experimentation to create a culture of learning through innovative environments, programs, and tools that help people nurture their curiosity about the world around them. This notion of experimentation is drawn from public visitation to museums and interaction with exhibits even while the exhibits are under development, and invites staff, visitors, teachers, youth, and the general public into this process of exploration to critique content, media, and science (Oppenheimer, 1976; Semper, 1998a). This philosophy, which first began with tinkering with exhibit prototypes on the museum floor, also pervades and influences Web and interactive media design (see Semper, 1998b) and the methods used to evaluate these designs. Because evaluation can sometimes be viewed as a stick rather than a carrot, we offer some alternative user feedback collection methods that are based on the overall philosophy and culture of the Exploratorium. Borrowing from a model of exhibit-design testing, these approaches to Web site evaluation engage visitors in a two-way dialogue as on-going legitimate participants providing feedback and input into the media development process.

In this paper, we draw examples from two National Science Foundation-funded informal science education projects, "Evidence" (http://www.exploratorium.edu/evidence )and "Mind" (http://www.mindinprogress.org), to illustrate how new evaluation approaches have been used to support Web design and development, and influence design changes.

## Case Study 1: Evaluating "Evidence" Web site Components

The goal of the Evidence Web site is to improve the public's understanding of a core concept of science, the nature of evidence, providing insight into how scientists know what they know, how they use scientific evidence to evaluate emerging theories and attempt to construct a functional understanding of the world by gathering evidence, and how they make discoveries based on evidence.

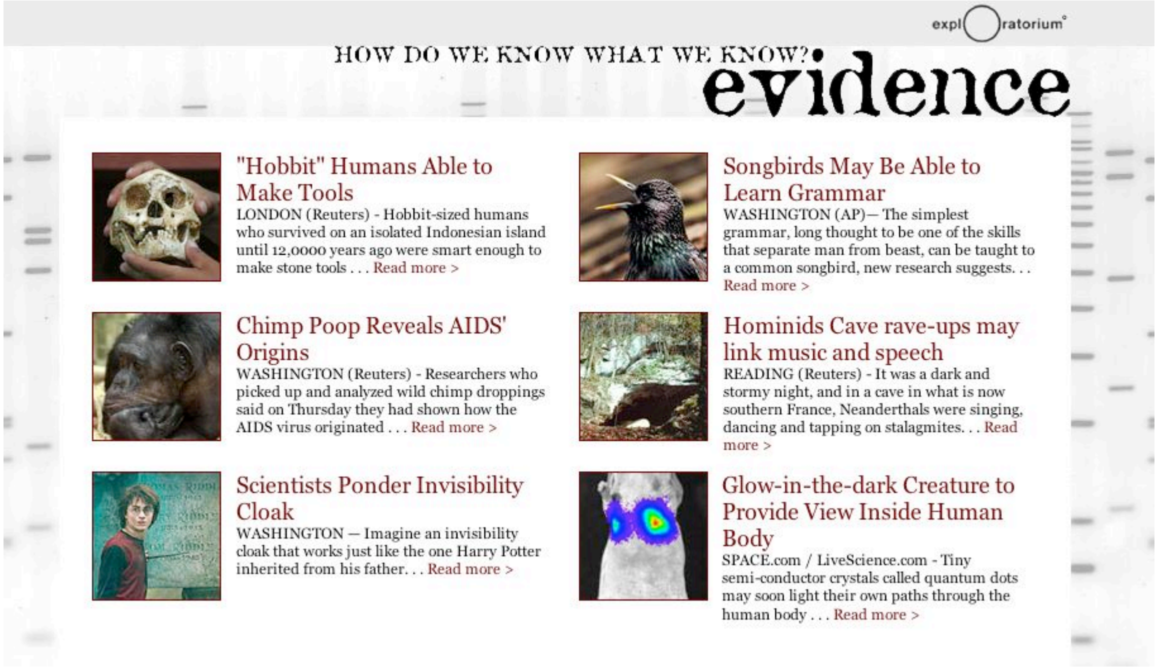
After conducting a front-end literature review to collect and review prior examples of how others have addressed the issue of scientific evidence, the next step was to hold a series of off-site meetings and design charettes with multiple stakeholders including Web developers, artists, writers, scientists, evaluators, and other staff to brainstorm topics. As ideas were brainstormed and sketched out, many open-design questions were posed: What might be an engaging way to organize and structure content for the on-line visitor? What questions would visitors have as starting points for inquiry? How can a site help provide a personally meaningful experience for the user? The results of the brainstorming process led to the emergence of four concept entry points for possible interactives and a site map: My Evidence, Evidence-To-Go, Ten Questions, and a Panorama of Scientists. In this case study, we elaborate on the design and evaluation of the Ten Questions concept. (My Evidence is elaborated on in a separate paper: Robinson et al., 2007.)

One assumption was that the structuring of Web content as a set of driving questions would be useful for visitors, both as a tool for exploring different topics on the site and as an instructional tool that could used in the future for visitors'own inquiries into new topics or scenarios. For the Ten Questions Web component, the following three categories of questions were identified:

1. Emergent curiosity questions: These are questions that the public raise (or would like answered) about the science and scientific evidence when they read science articles, and that emerge naturally because they are personally relevant and meaningful to the individual exploring the site.
2. Instructional scaffolds: These are instructionally designed questions that people should be asking themselves when reading science articles in order to become more critical thinkers regarding science, scientific evidence, and scientific knowledge.

3. Expert guided questions: These are questions that scientists commonly ask themselves when critically reviewing evidence, science, or scientific knowledge.

Thus, the goal of the evaluation was to first identify the top ten emergent curiosity questions that the public would be likely to ask or would like answered when presented with different science evidence in the context of the Web site for the Ten Questions concept. Various layouts for Ten Questions led up to the development of a test site that involved gathering rapid feedback by grabbing a handy sample of friends and staff to provide "over-the-shoulder" critiques. Because news articles provided a simple way to present evidence in a familiar and accessible way to the public, eight on-line news articles were selected from Internet news sources with varying levels of proven reliability in reporting accurate facts, and then framed into a "toss-away" test site (Figure 1).



"Hobbit" Humans Able to Make Tools
LONDON (Reuters) - Hobbit-sized humans who survived on an isolated Indonesian island until 12,0000 years ago were smart enough to make stone tools . . . Read more >

Songbirds May Be Able to Learn Grammar
WASHINGTON (AP)— The simplest grammar, long thought to be one of the skills that separate man from beast, can be taught to a common songbird, new research suggests. . . Read more >

Chimp Poop Reveals AIDS' Origins
WASHINGTON (Reuters) - Researchers who picked up and analyzed wild chimp droppings said on Thursday they had shown how the AIDS virus originated . . . Read more >

Hominids Cave rave-ups may link music and speech
READING (Reuters) - It was a dark and stormy night, and in a cave in what is now southern France, Neanderthals were singing, dancing and tapping on stalagmites. . . Read more >

Scientists Ponder Invisibility Cloak
WASHINGTON — Imagine an invisibility cloak that works just like the one Harry Potter inherited from his father. . . Read more >

Glow-in-the-dark Creature to Provide View Inside Human Body
SPACE.com / LiveScience.com - Tiny semi-conductor crystals called quantum dots may soon light their own paths through the human body . . . Read more >

Fig 1: A "toss-away" test site created for the purposes of eliciting visitors' top ten questions about scientific evidence.

## Evaluation by floor recruitment

One Web developer and two evaluators (internal staff and external contractor) worked as a team to recruit visitors directly from the museum floor to evaluate the test site. Testing took place on one weekday and one weekend. Signs were placed by the entrance of a cordoned-off space using stanchions to create an "evaluation corral" near high traffic museum exhibits (Figure 2). This corral contained small tables on wheels with wireless laptops, computer mouses, and headphones. The team requested that visitors read at least two of the six articles and encouraged them to take as much time as they wanted (visitors spent anywhere between nine to twenty-three minutes reading). The team also set a stack of scratch paper and a pen next to the computer and prompted visitors to write down any questions they wanted answered about the articles – specifically, about the evidence presented, as follows:

> "While you are reading these articles, please write down any questions that come to mind that you are curious about and would like answered, particularly relating to the scientific evidence presented."

While visitors were seated at the computer, the evaluator observed their movement as they read through the articles. When visitors indicated they were done reading the articles, they were escorted by a member of the evaluation team to the interview table. The evaluator and the developer took turns first asking the visitors to restate and elaborate on the questions they had about the articles, using the following questions and probes:

> "Can you tell me about some of the questions that came to mind when you were reviewing the articles? Was there any evidence presented in the article that raised your curiosity? Why did this (topic, article, piece of evidence) raise questions for you?"

The evaluator next showed visitors a list of seventeen possible questions about scientific evidence, and asked them to check off the key questions they had about scientific evidence. The question order was rotated to counteract a possible order effect.

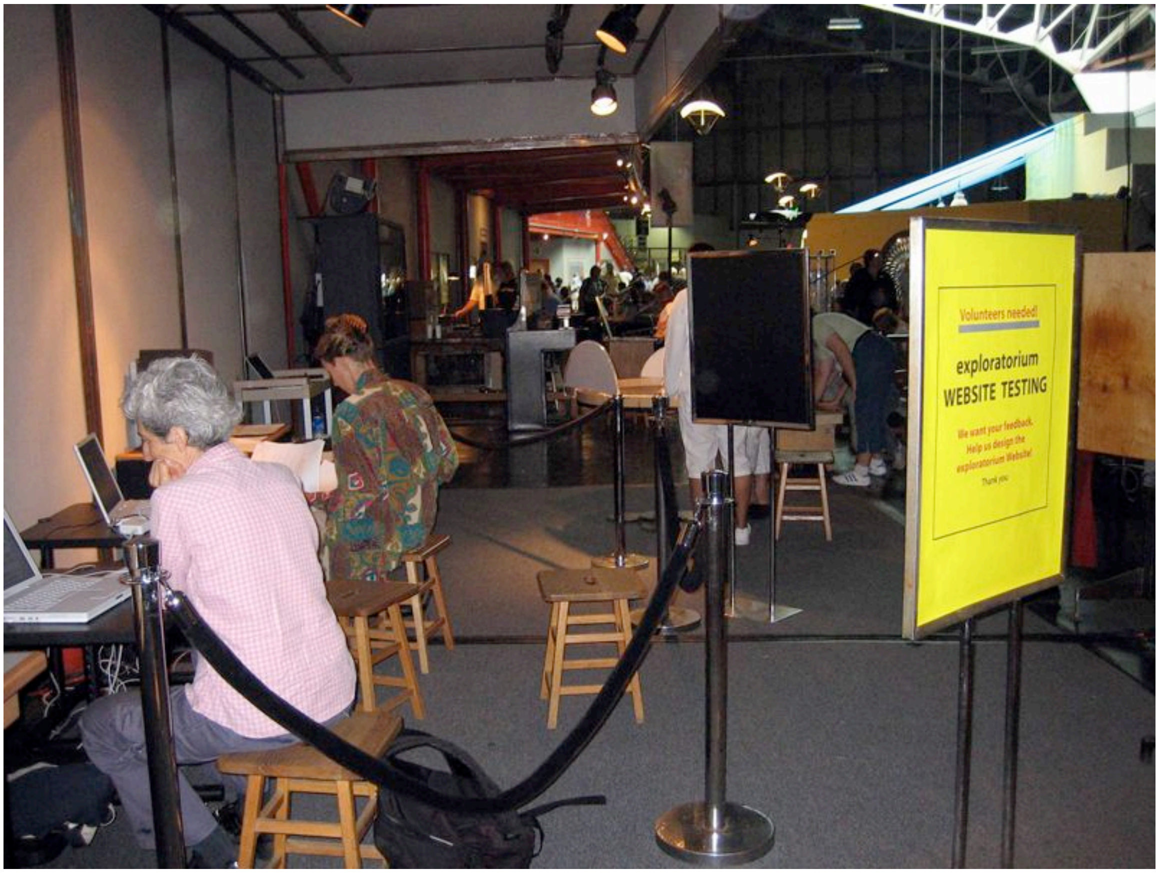


Fig 2: A transient evaluation corral set up in an exhibit space at the Exploratorium that was used to recruit visitor and test design ideas and Web prototypes.

Finally, visitors were given an opportunity to share any other comments they had about their experience reading the articles and reflecting on the evidence presented. At the conclusion of the interview, the evaluator thanked the visitors for participating and gave them a small gift from the Exploratorium Store. The results yielded a core set of questions (Table 1) that were then further refined as questions to be used in the next iteration of the Web site design.

### Lessons learned

While a more elaborate evaluation scheme was possible using a test site with custom Javascript to randomize posting of the questions, capturing user screen clicks to mark questions and evaluating user behaviors, or using videotape data of visitors' talking aloud, we operated on the evaluation principle of "fast, cheap, and easy" in the spirit of other "discount" usability evaluation methods (Nielsen, 1994a).

The evaluation enabled us to quickly learn about the amount of time visitors spent on the computer, the articles they most often selected, their interest in the articles, the emergent questions they had about the evidence presented in the articles, and broader questions about scientific evidence that developed as they read the articles.

Table 1: Visitors' top ten questions about evidence after reading the articles

| Ten Questions |
|---|
| Q1: Where is this source of information coming from? |
| Q2: Has their work been verified by another source or by another scientist? |
| Q3: Are the methods or methodologies sound? Are they explained? |
| Q4: What other sources or reports are there on this topic? |
| Q5: Does the report indicate the limits of their claim? |
| Q6: Is this evidence confirming an existing theory, challenging an existing theory, or reporting on something for which there is no theory? |
| Q7: What makes science wrong? Who decides, and how? |
| Q8: How was it discovered? |
| Q9: Is this evidence related to something I think I know already? |
| Q10: Is the data presented in a biased or neutral way? |

To be able to share these kinds of findings with the project team, we then prepared an evaluation report that summarized everything we had learned from observing the visitors as they read the articles as well as reviewing and sorting their written and verbal responses to our questions. We also looked at the above findings from multiple standpoints. For example, in addition to sorting visitors' questions according to the articles to which they referred, we also sorted their questions based on the kinds of issues they raised about scientific evidence more broadly. Finally, in looking across the various data sources we had compiled, we examined more closely the themes that emerged from visitors' top question picks (from the 17-item checklist we provided) in the context of their written questions, drafted while reading the articles, and their comments in the follow-up interviews (when they were given an opportunity to elaborate on their questions). By triangulating the data in this way, we observed that visitors' questions about scientific evidence could be categorized into about nine distinct themes. The development team then discussed each of these themes, and incorporated elements of visitors' thinking (and their use of language) about scientific evidence to help inform the next phase of prototype development.

## Summary of evaluation approach

The Web site components are tested formatively and iteratively in the same way exhibits are tinkered with and designed at the Exploratorium: by pushing a design on the museum floor and grabbing handy samples of people (visitors, friends, staff) to provide immediate and concrete feedback.

Rather than fully develop a Web site or even a fleshed-out Web prototype, intermediate Web site "toss-away" components are developed for the sole purpose of helping to test a concept and answer a design question.

The developer and evaluator partner together to make first-hand and shared observations of visitors using the interactive. They also take turns interviewing users. By having multiple simultaneous viewpoints, the team can better come to consensus on the interpretation of data, observations, and the next steps needed for the design.

We make the tenuous assumption that the use by visitors to the Exploratorium is similar to the behavior of remote users who use the Web site. A potential limitation of this approach is that the museum floor is an imperfect proxy for authentic user behavior and user intent in an off-site setting. However, for purposes of concept testing and basic usability testing, we have found this approach to be useful.

## Case Study 2: Methods in the "Mind" Project and the "Mind" Web site

"Mind: Attention, Emotion, and Judgment: How do minds figure out what to do?" is a development project funded by the National Science Foundation to create a new collection of exhibits and experiences that allow visitors to explore and experiment with the workings of their own minds. As part of this project, the Exploratorium is developing a Web site that provides interactive experiences that allow users to explore aspects of their own attention, emotion, and judgment. Many of these interactives include activities in which users can add their own input to a larger database for further analysis and interpretation.

### Evaluation plan

Our evaluation plan is separated into two broad stages synchronous with the stages of Web development.

#### *Evaluating individual interactives*

Development of the "Mind" Web site begins with the development of different interactive media pieces, and early formative evaluation has consequently been focused on the individual interactives, or Web components. In this stage, we were primarily interested in:

- Collecting enough data to determine if the interactives reveal larger patterns in participants' reactions. This was particularly important for components modeled after classic psychology experiments that rely on showing a statistical difference between two treatment groups,
- Identifying usability problems for each interactive,
- Gauging users' difficulties in interpreting their own, individual experience and the collective dataset.

### *Evaluating the overall site*

The overall site will then be designed to integrate the different interactives into a coherent experience for our Web audience. It is during this stage of development that we will:

- More clearly articulate the Web site's potential audience, including any particular group that has a special interest in this subject area,
- Identify difficulties understanding the overall message of the Web site,
- Determine and address navigation and organizational issues with the overall site,
- Conduct more in-depth usability studies with later stage prototypes.

Our work to date has been on individual interactives, and the remainder of this paper will focus on describing the procedures and tools we have developed for doing so as well as discussing the advantages and limitations of each in collecting information that informs our Web development. In the later stages of development, we anticipate using additional Web evaluation methods, including cognitive walk-throughs and usability testing with beta-testers recruited from specified populations. These tests will be more concerned with audience and context of use.

## Evaluating Web components at the museum

### *Involving other experts*

We relied on a set of processes and tools to help us evaluate the individual Web components. First, we tried to involve people with other expertise early and often in the development process. That is, we engaged writers, content experts and scientists, and other developers in a form of heuristic evaluation (Nielsen, 1994b), wherein each person was asked to use the interactive separately. Then all the participants gathered to critique the interactive as a group. This was one of the ways we began to collect information about usability and comprehension issues.

### *Adapting physical exhibits to Web versions*

The Web developer and evaluator on the project also worked collaboratively with the exhibit developers to identify physical exhibits that may work better in the virtual realm. Observations we gathered for the physical exhibit then informed the development of the Web component. For example, an exhibit called Magnetic Faces was adapted to the Web after we discovered that visitors wanted help identifying the different magnetic parts that made up a facial expression, something easy to provide in the Web version but more difficult in the physical version.

### *Using a floor kiosk*

Our main method of collecting early user-feedback involved using the Mind On-line Interactive Kiosk (MOIK) with visitors on the Exploratorium floor. MOIK is a computer configured to run in kiosk mode (i.e., Firefox plus R-kiosk extension) that makes available a set of Web components in a version that is more appropriate for floor use (e.g., fewer choices, bigger buttons, bigger text than its Web counterpart).

MOIK has a wireless Internet connection, a secure keyboard (i.e., a cheap keyboard with glued-down function keys), a trackball, a touch screen, and room for USB peripherals (Figure 3). A Web site with the Web components to be evaluated runs on MOIK and is reconfigurable by computers with an Internet connection. This allows changes — by the Web developer, evaluators, or a scheduled computer script — to be easily made during development.

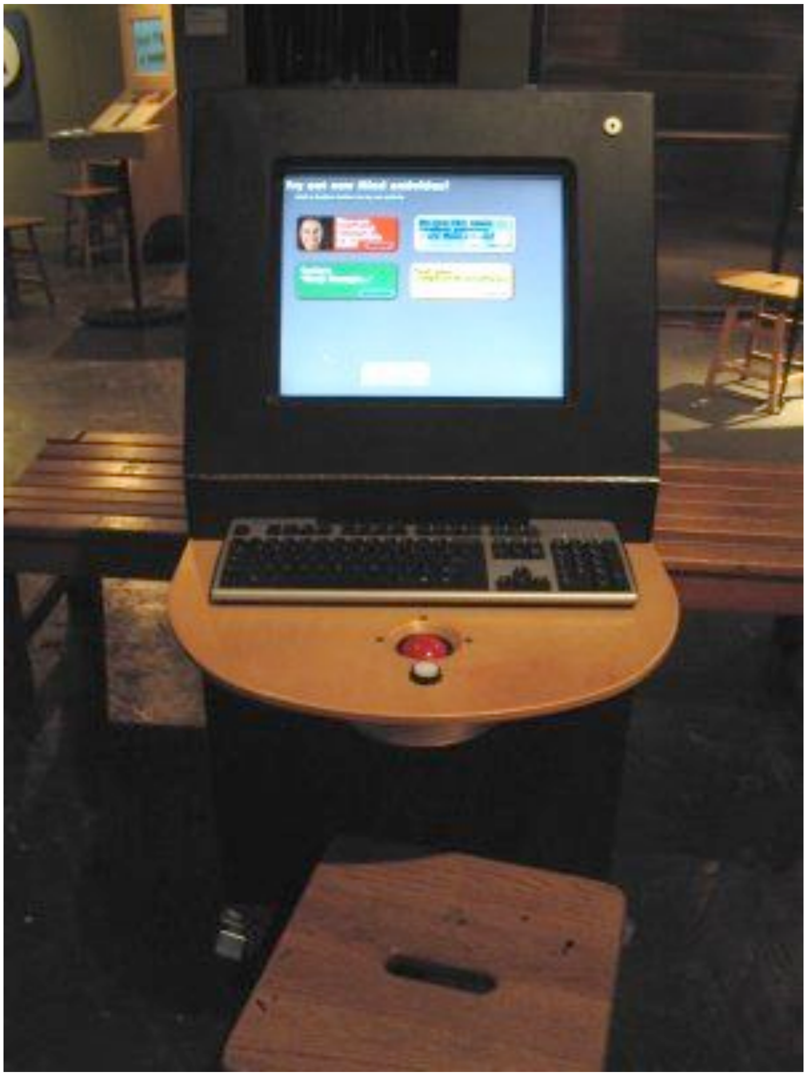


Fig. 3: The MOIK (Mind On-line Interactive Kiosk)

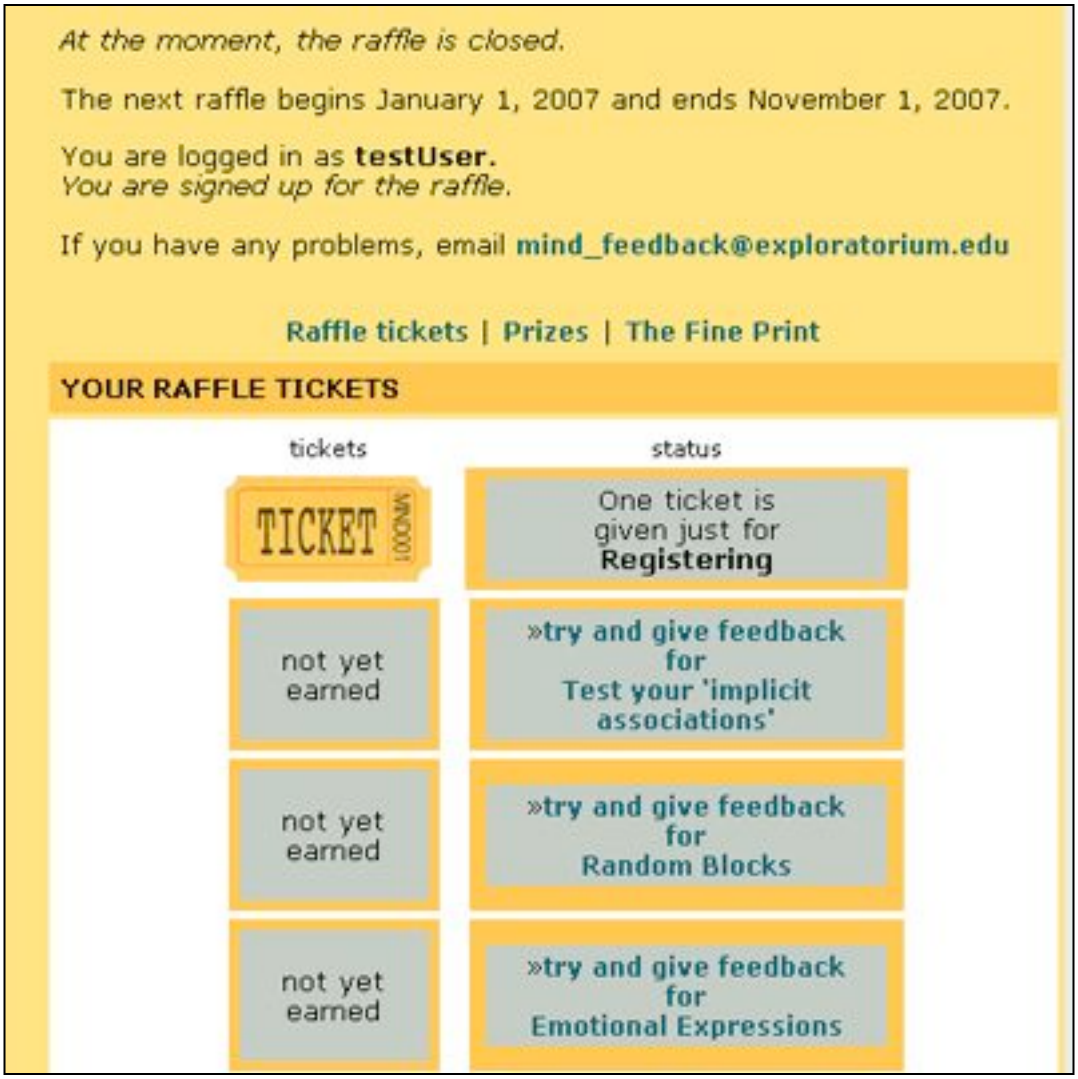
Meanwhile, a simple tracking script in PHP and MYSQL that logs every screen jump runs on each Flash interactive and Web page. This click-stream data gives us an indication of which interactives are being used and which pages of each interactive are visited.

## Advantages

### *Rapid prototyping of aggregated data*

MOIK provides us with several advantages in collecting informative data for our Web interactives. First and foremost, it allows us to begin to collect initial data about our Web components without building a more expensive beta site. A beta site would require developing more formally polished interactives, recruiting a user base, developing the overall Web site that introduces and integrates the interactives, and then proceeding with regular Web site evaluation to obtain a handful of uses.

This advantage was not immediately obvious. In early development, we created a login-based Web site, complete with an on-line raffle using digital tickets that would be automatically collected if a logged-in user completed an activity, or submitted feedback responses to surveys (Figure 4). While museum staff participated in the raffle, the site was too premature to merit dispersion into cyberspace. Reasons for this included needing a "home base" URL for dispersed media experiments that would persist for several years, orienting Web site newcomers to the project while the Web project itself was still being defined, setting up a more reliable login database, and attracting diverse groups of beta testers. Because it became clear early on that staff use and testing in early development was not an adequate replacement for heavier traffic and that it was going to be less bang for the buck to have on-line testing, we conceived of and developed the MOIK to provide more immediate visitor feedback for new interactives. In six months, the kiosk logged more than 200,000 interactions, a number much larger than we would ever have expected for a beta site in its early stages.



*At the moment, the raffle is closed.*

The next raffle begins January 1, 2007 and ends November 1, 2007.

You are logged in as **testUser.**
*You are signed up for the raffle.*

If you have any problems, email **mind_feedback@exploratorium.edu**

**Raffle tickets | Prizes | The Fine Print**

**YOUR RAFFLE TICKETS**

| tickets | status |
|---|---|
| TICKET 100001 | One ticket is given just for **Registering** |
| not yet earned | **»try and give feedback for Test your 'implicit associations'** |
| not yet earned | **»try and give feedback for Random Blocks** |
| not yet earned | **»try and give feedback for Emotional Expressions** |

Fig. 4: Screen showing the on-line raffle using digital tickets that would be automatically collected if a logged-in user completed an activity or submitted feedback responses to surveys.

This ability to reach a large set of users quickly was particularly important for testing psychology experiments, relevant to the developing "Mind" exhibition. Many psychology experiments depend on comparing responses from two independent groups. This requires a decent sample size from each group in order to see statistical differences in user responses. The MOIK expedites this process by allowing us to collect the data quickly and therefore determine if an interactive modeled after a lab experiment is robust enough to show results in a less controlled environment with more diverse "subjects."

Additionally, this ability to rapidly prototype aggregated data interactives has allowed the developer to begin to isolate major design issues. Key among these issues are the amount of text to read, problems interpreting and designing statistical data, conveying distinctions between two comparative groups, locating the user's data point in relation to that of the group, and needing many screens to convey all the information. Thus far, this has encouraged exploring a different approach: using simple, immediate experiments, less explanatory text, and giving the data output significantly more graphic design attention. This efficiency promises to be a great asset as the museum world looks to aggregated data for Web and museum visitor exhibits.

### *Real-time observation*

Because the MOIK is on the museum floor, we can observe directly what people do with these Web interactives. Informally, a Web developer can watch visitors struggle and succeed with their new interactives. Formally, we have made systematic observations of visitors using the MOIK, noting who stops, when they stop, what interactives they use, what if anything they say when they're at the MOIK, and when they leave the kiosk. We have also used an overhead Web cam to disentangle different visitors' experiences at the MOIK. These observations are coordinated with the click-stream data to identify which interactives visitors use and which parts of each Web component they look at (Figure 5) because click-stream data alone can't reliably tell us when one visitor has left and another has begun to use the MOIK.

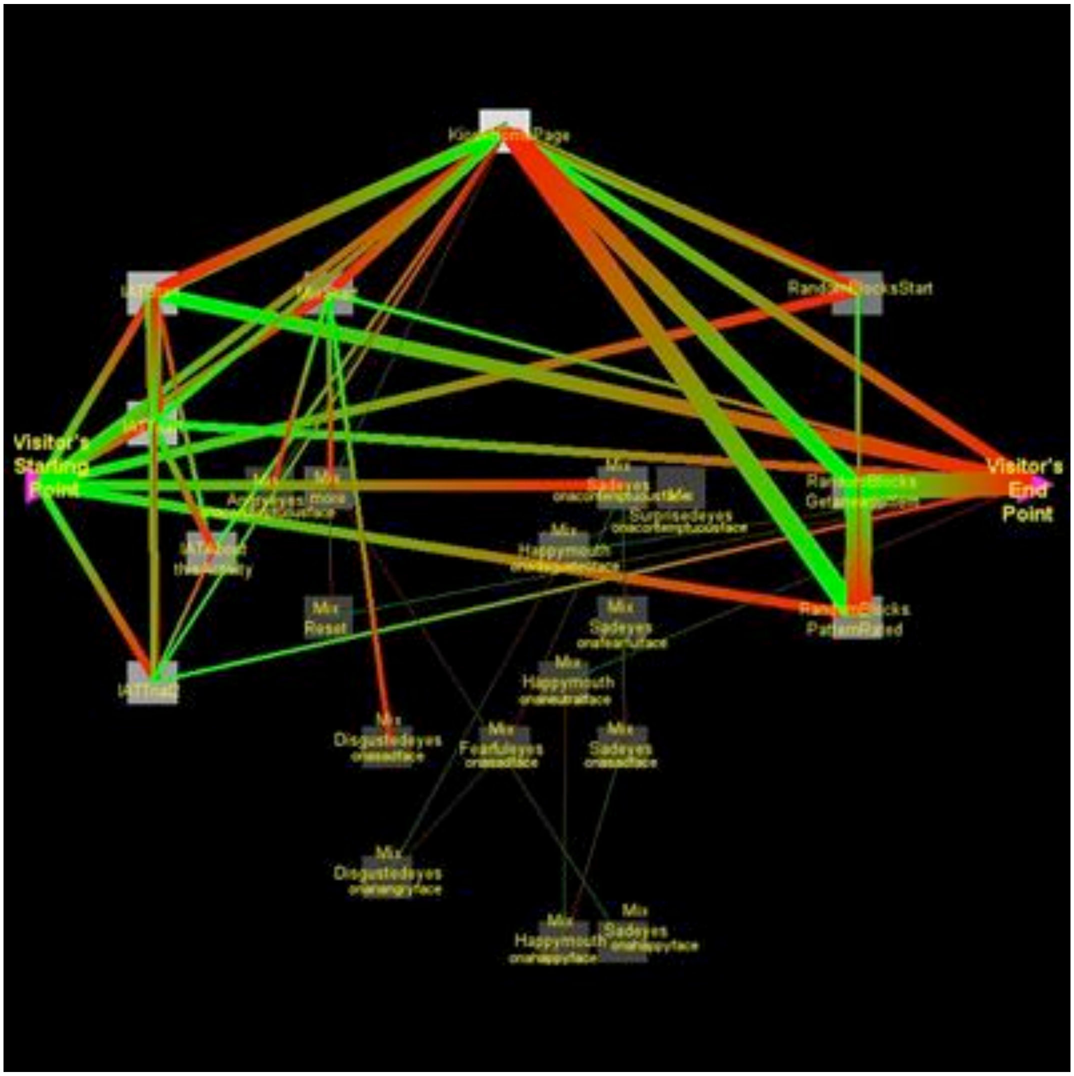

Visitor's Starting Point
Visitor's End Point
RandomBlocksStart
RandomBlocks PatternRated
IATTrial2
Mix Reset
Mix Sadeyes onacontemptuousface
Mix Surprisedeyes onacontemptuousface
Mix Happymouth
Mix Sadeyes onafearfulface
Mix Happymouth onaneutralface
Mix Disgustedeyes onasadface
Mix Fearfuleyes onasadface
Mix Sadeyes onasadface
Mix Disgustedeyes onaangryface
Mix Happymouth onahappyface
Mix Sadeyes onahappyface

Fig 5: Web components that floor visitors looked at on the MOIK. This is from click-stream data logged by the MOIK, sampling one visitor every 10 minutes from 10 a.m. to 3 p.m., one day in December. Each box represents a page. The brightness of a box indicates the number of visits that page received. The thickness of a line between two boxes indicates the number of times visitors jumped between those two pages.

We also have the advantage of administering an interview immediately after a visitor has finished using the MOIK. These interviews help identify usability issues and comprehension issues in addition to those issues brought out by experts in our heuristic evaluation. (See Ma, 2006, for an example.)

### *Qualitative data collection*

In addition to collecting quantitative data, we are also using the MOIK to collect more qualitative data to help us generate additional exhibit development ideas and provide new insights. Specifically, we designed an interactive, “Today's Deep Thought,” that asks visitors to respond to philosophical questions we have authored. We try to keep these questions engaging and thought-provoking for visitors, and we can obtain 20 to 80 responses from visitors each day. We convey to visitors that they’re helping to develop new exhibits, and we don’t share visitors’ responses. This means that we don’t have to devote resources to moderating content, which usually contains a mixture of swearing, key-pounding, and thoughtful responses. These visitors’ contributions give us a rough idea of how they think about certain questions, which can bring up alternative viewpoints outside of the development team’s collective knowledge. We have been able to experiment with different wording of the questions to gauge what encourages visitors to write more thoughtful answers, as well as to experiment with the integration of graphics to obtain responses from younger audiences.

## Limitations

Although the MOIK provides us with several advantages in formative evaluation, there are limitations to the data that we collect, key among these is that the data reflect the use of the interactives on *the museum floor*. How a person uses these interactives on a computer on the Web can be different from his or her use at the Exploratorium’s kiosk.

One factor affecting interactives use is that a majority of people who visit the Exploratorium come in groups. Their interactions at the MOIK are therefore a social experience that is likely to be different from the typical experience at a Web site visited by an individual, not by a group gathered around a shared computer monitor. In fact, in one of our observations, we found that 19 out of 21 visitors who stopped at the MOIK came to the kiosk with another person. Different members of a group may be helping others in that group; that is, reading aloud, encouraging engagement, or interpreting what’s happening. While these interactions may be different from the typical Web interaction, they may also hold some hidden benefit for visitors who may be listening to these conversations.

Another factor to consider is that the museum can be a dynamic and sometimes chaotic place. Some of the Web interactives require focus, concentration, and time for reflection. These interactives fare much worse on the Exploratorium floor than in their eventual setting. For example, we have found it difficult to replicate the results of some classic laboratory experiments in psychology with our interactives based on those experiments. However, because the museum floor environment is so demanding, it forces the developers to sharpen the design of the interactive in ways that are most likely to also sharpen the home experience by addressing visual design, by use of graphics and illustration, by simplifying text, and by paring down to essential content goals.

Finally, as was the case with visitors who tested the “Evidence” Web components, it isn’t clear if the people who stop at the kiosk are representative of the people who would visit the Web site. At the early stages of development and evaluation, this is not as critical a concern because we are still interested in having our interactives work for as broad as an audience as possible. However, this points to the importance of supplementing these evaluations with studies using the actual Web site with actual users. Nonetheless, using a kiosk on the floor has been invaluable in allowing us to collect early data on how users interact with our Web components.

## Conclusions

As we hope these examples have shown, Web design can benefit from, and have a tight interrelationship with, the physical museum and with evaluation. An essential feature of this process is the collaboration between Web developers and evaluators. As exhibits and museum experiences become more participatory and involve visitors in the voice, messages, and construction of a shared experience or exhibit, this pushes on the Web 2.0 concept that visitors are both contributors, authors, and critics of museum designs, whether on-line or on the floor. As museum professionals, we can capitalize on the museum opportunities to generate new knowledge about user experiences for the Web through these new forms of formative evaluation.

## References

Ma, J. (2006). *Rack Your Brain on the Mind Online Interactive Kiosk.* Retrieved January 15, 2007, from Exploratorium, Visitor Research and Evaluation Web site: http://www.exploratorium.edu/partner/pdf/moik_rp_02.pdf

Nielsen, J. (1994a) *Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier*. Retrieved January 15, 2007 from Jakob Nielsen's Web site: http://www.useit.com/papers/guerrilla_hci.html

Nielsen, J. (1994b) Heuristic Evaluation. In *Usability Inspection Methods*, Jakob Nielsen and Robert L. Mack, Eds. John Wiley & Sons, Inc., pp. 25-62.

Oppenheimer, F. (1976) *Everyone is You or Me.* Reprint from Technology Review. Alumni Association of the Massachusetts Institute of Technology. Volume 78, Number 7. Retrieved January 15, 2007 from: http://www.exploratorium.edu/frank/everyone/everyone.pdf

Robinson, L. Beck, D., Tesler, P. (2007) My Evidence: Who's the Authority Here? To appear in the Museums and the Web 2007 Conference Proceedings. San Francisco: Archives and Museum Informatics.

Semper, R. (1998a) Bringing authentic museum experiences to the Web. Museums and the Web 1998 Conference Proceedings. Pittsburgh: Archives and Museum Informatics. Retrieved January 15, 2007: http://www.archimuse.com/mw98/papers/semper/semper_paper.html

Semper, R. (1998b) Designing Hybrid Environments: Integrating Media into Exhibition Space in the Virtual and the Real: Media in the Museum, Selma Thomas and Ann. Mintz, Eds., Washington, D. C.: American Association of Museums, pp. 119-127

## Acknowledgements

We would like to gratefully acknowledge the contributions from evaluator Valerie Knight-Williams and the editorial expertise and assistance from Ruth Brown, Ellyn Hament, and Laura Jacoby at the Exploratorium. The authors also acknowledge the National Science Foundation support provided under grants ESI-#0452128 and ESI- #0307927.