The Outdoor Exploratorium:
Experiments in Noticing and Understanding

A Summative Evaluation of the First Seven OE Exhibits

by Beverly Serrell, Serrell & Associates, Chicago, IL

March 2009

Table of Contents

INTRODUCTION

In 2001, the Exploratorium was awarded a $1.3 million grant from the National Science Foundation (NSF) to create a new group of pedagogical tools to help visitors notice and explore complex natural phenomena. Instead of simulating air or water movement indoors, the Outdoor Exploratorium (OE) exhibits would allow direct experience of the dynamic natural environment, such as heat, wind, waves, and sounds. Envisioned as a group of 20 to 25 interactives located in a 10,000-square-foot area adjacent to the Palace of Fine Arts (PFA), where the Exploratorium is housed, the outdoor exhibits would reinforce the phenomena-based experiences that visitors would find indoors and vice versa.

Changes and delays in the Exploratorium's renovation and relocation plans, along with project personnel reassignments, caused the OE team to reconceive the group of exhibits to be disbursed as stand-alone units installed at Fort Mason, a historic site at the Golden Gate National Recreation Area (part of the National Park Service), located about a mile from the Exploratorium. This site and partnership created new challenges for the development and evaluation of OE exhibits, and in the fall of 2008, as the project neared the end of the grant period (February 2009), seven of the 20 exhibits were ready for summative evaluation.

The exhibits reviewed in this report are

- Architectural Mirage
- Speed of Sound
- Corrosion Wedge
- Wave Oscilloscope
- Bridge Thermometer
- Wind Arrows
- Audio Posts

A brief description of the intended visitor experience for each exhibit will be included in the "Findings" section of this document.

As the summative progressed, it became clear that a subsequent study would need to be pursued--one that looked at the total group of exhibits at Fort Mason, not just a subset, and NSF approved a no-cost extension of the project until Aug. 31, 2009. By the second summative, all the exhibits will have undergone at least one iteration of formative evaluation. The findings from this first summative (1) present a snapshot of the seven stand-alone exhibits, (2) provide remediation for those exhibits, and (3) otherwise serve to inform the development of the remaining exhibits and overall collection.

GOALS

A primary intended output of the project was to create exhibits that foster a sense of creative investigation and appreciation of both subtle and powerful elements in the outdoor world. The visitor outcome goals of the OE project, listed in the NSF proposal, were to

- help visitors develop skills in noticing natural phenomena outdoors
- help visitors explore complex systems and interactions at play in an outdoor environment
- help visitors come to a deeper understanding of the phenomena by applying scientific concepts and principles to the outdoor environment

The summative evaluation goals were to assess the degree to which the exhibits achieved the project's goals and to link the evaluation findings to the NSF's *Framework for Evaluating Impacts of Informal Science Education Projects*, which includes five categories of evidence for impacts (visitor outcomes) that are related to learning: knowledge, engagement, attitude, skills, and behavior. The skills category became the primary focus of the summative evaluation because it related most directly to the objectives and affordances of the OE exhibits under study.

OE exhibit developers defined “noticing skills” as visitor behaviors (what people did and what they said) that provided evidence for a wide variety of desired outcomes, including

- *Tracking,* or *transferring* site-based noticing to or from other environments
- *Recognizing patterns* in flow or repeating events
- *Identifying interrelationships* between forces or objects, not isolated events
- *Making comparisons* or *correlating*, looking for observable links between phenomena
- *Venturing a hypothesis* for future noticing, or *probing*, asking a question to direct further noticing and to uncover new observations or insights
- *Valuing* the mundane, ordinary, or naive in a new way
- *Gaining precision* or improved noticing of details, paying attention more closely
- *Translating,* or using a tool to make what is difficult to notice “noticeable”
- *Gaining motivation* through rewards from paying attention more closely, reinforcing an intention to repeat
- *Challenging assumptions* through revised observation or expectation

Noticing skills go beyond casual observation. Evidence for them usually results from sustained attention and is complex because it includes and incorporates visitors’ prior knowledge, attitudes, and what they found meaningful in using the exhibits. The types of noticing skills listed above are overlapping and not exclusive. They are not listed in a specific order, nor do they have a hierarchy.

As part of the preparation for this summative study, the exhibit developers listed the specific behaviors and outcomes they hoped to see and hear for their own exhibits. These outcomes were more concrete statements that supported the overall goals and the noticing skills listed above, and they provided a reference for comparing what visitors said about the exhibits during the evaluation.

Each exhibit developer brought his or her own individual expertise to bear on the definition of noticing skills. He or she then participated with the evaluator in the analysis of visitor comments to look for evidence of the skills in his or her exhibit(s). The criterion was, “Did this comment give you goose bumps of joy when you heard it?” That is, did it strongly resonate with your objectives? Thus, the exhibit developers shared responsibility for defining success. Given the somewhat elusive nature of defining and identifying noticing outcomes, the collaborative efforts with the evaluator in this way helped reinforce her efforts. The examples (quotations) used to illustrate positive evidence of impacts in the “Findings” section were selected by consensus.

Other overall output goals for the exhibits--the characteristics and qualities they all should have--developed during the exhibit design stages and were used to keep the Fort Mason partners informed of the Exploratorium’s intention. These included fundamental appeal; interesting form and content; participatory activity; elements that stimulate curiosity and lead to broader questions; innovation and originality; balance of content and aesthetics; and alignment with and enhancement of the site. The aspects of stimulating curiosity and evoking broader questions resonated with the visitor outcomes of noticing skills and will be included as part of them. The rest will be considered separately as to how they applied to each exhibit and/or as a group.

METHODS

Cued interviews were used to assess the effectiveness of the seven OE exhibits installed at Fort Mason. Visitors were recruited by professional data collectors to look at an exhibit, after which they were asked a series of demographic and open-ended questions (see form in Appendix). Cuing and the use of small samples were partly in response to a National Park Service requirement regarding social science research on NPS property. A sandwich-board sign announcing the Exploratorium’s exhibits and inviting participants to try them out was on display during data-collection periods.

Data collectors transcribed the answers for each participant at each exhibit for an overall total of 81 samples (10 to 14 visitor samples per exhibit) over 15 days of data gathering (afternoons of Friday, Saturday, or Sunday) during October and November 2008.

Transcriptions were reviewed by the evaluator (Beverly Serrell) and the Exploratorium exhibit developers (Bryan Connell, Shawn Lani, Charles Sowers, Maz Kattuah, and Ulrika Andersson) for three basic categories:

- Evidence of noticing skills, i.e., things people said that resonated with the intended visitor outcomes listed above.
- Visitors' compliments, opinions, suggestions, misunderstandings, or complaints about the exhibits.
- Other feedback that fell into the category of simple noticing or casual observation. Not "goose-bump" comments.

Due to the overlapping nature of the noticing skills, coding the transcriptions beyond those mentioned above was not attempted. In a preliminary test of the questionnaire/interview instrument, the feedback from visitors could not easily be distinguished between "knowledge" and "noticing skills," but the method clearly captured indications that visitors were broadening their immediate experiences of noticing the targeted phenomenon. As one exhibit developer commented about the fact that visitors' feedback resonated well with the exhibit's intentions, "I'd say those are heartening results. I'm not sure how much closer we're going to get."

LIMITATIONS

Usually summative studies involve more than one assessment method, but in this case the restrictions of the location and the timing prevented multiple methods and caused challenges for the process. Unobtrusive observations were not a practical choice: Some of the exhibits were installed in low-traffic areas, and the design of the exhibits was meant to blend with the environment. Few visitors to Fort Mason would be likely to wander past some of the exhibits, and many people would not even realize there was an exhibit to look at.

There were other challenges to collecting data:

• The data-collection period corresponded to the installation of the exhibits in October and November, a time when the weather at the bayside Fort Mason was increasingly cold, foggy, windy, or otherwise miserable for standing outside for extended periods.

• Some of the exhibits had weather-dependent phenomena (heat, wind, fog), so all of the exhibits did not necessarily "work" all the time.

• Fort Mason is a site for many public events, which involved changes in the accessibility of the exhibit areas. Trucks and other equipment were moved in and out for the events (e.g., antiques, snow boarding, art, air shows). Many visitors were on their way to or from an event, or were not expecting to encounter exhibits outdoors, and they did not want to be interviewed.

The refusal rate was unusually high. Some days more than 50% of the people declined to participate. Nevertheless, the persistently friendly and adaptable data collectors were able to recruit and interview people who turned out to be, for the most part, eager to help and nice.

In fact, the people who participated in the interviews were almost *too* nice. They represented the *creme de la creme* of willing, cooperative, helpful, hard-working, eager-to-please subjects. The degree of "courtesy bias"--the tendency for respondents to give answers that they think the interviewer wants to hear, rather than what they really feel--was high. Taking into account the unusually high number of refusals, those who did participate gave far more positive feedback than could be expected from a random uncued population of subjects with a lower refusal rate. The findings must be considered in this light. The limitation is the potential lack of negative feedback. People who opted out may have already perceived the exhibits as uninteresting or confusing.

Another limitation to this summative evaluation was that the proposed extent of formative evaluation to be done on the OE exhibits was not accomplished. Three of the exhibits had not been prototyped, and none of the interpretive graphics had been tried out. The four exhibits that were prototyped had been tested in substantially different forms from the ones evaluated during the summative; they had not benefitted from multiple iterations of revisions, which can often improve the final success of interactive exhibits.

Research on visitors to National Park Service sites is understandably restricted, and methods that would involve giving visitors questionnaires or collecting personal data (e.g., for follow-up questions) were not allowed for this study. There were other limitations to the methods imposed by the NPS for using NPS visitors as study subjects: Visitors had to be informed of the purpose and participate on a voluntary and anonymous basis; questions had to be open-ended; interventions needed to be brief and have as little impact as possible on as few people as possible. Aside from being limited to small sample sizes--to keep the "public burden" to a minimum--these restrictions did not prevent us from getting a lot of useful data in a short amount of time.

FINDINGS

The first part of this section reviews the findings that assess the success of the visitor outcomes. The findings include information about the demographics of the sample (gender, ages, group size, interests), the amount of time spent with the exhibit, answers to the open-ended questions, and the data collectors' observations and notes. Information that can be summarized for the whole sample (N = 81) will be reported first. In the second part, each exhibit will be reviewed separately. Visitor feedback through cued interviews was also used to review the success of the output goals--the characteristics and qualities of the exhibits--but in a more anecdotal than quantitative way, and this will be discussed in a later section.

Demographics

There were fewer family groups (adults and children), more males, and more adults by themselves in this summative study, compared to the usual evaluation sample of visitors at the Exploratorium. Of the total sample

- 89% were adults only, the rest were adults with children in the group
- 43% were singletons; 35% in groups of two, 12% in groups of 3 people
- 85% were adults; 12% were seniors. No children were interviewed

Due partly to the often scant number of potential visitors, Fort Mason employees were included a few times in the testing.

Special Interest

Interviewees identified their special interests, knowledge, or training in subjects related to the exhibits as having education in physics, science, art, maritime, architecture, sailing, building, and welding. But most (84%) said they had no special interest.

Time Spent

Overall, most visitors spent a minute or two interacting with the exhibits and the interpretive graphics. In one case, Wind Arrows, most visitors spent less than that. In another, Mirage, most of the sample spent a longer amount of time. These exceptions will be discussed under the individual exhibit reports.

General Issues

Several exhibits drew comments on their unobtrusive nature. Visitors said that if the data collector had not pointed out an exhibit, they would not have seen it. This was seen as both a positive and a negative aspect for different people. For example:

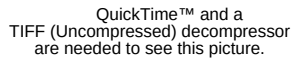
It's elegant. It's not where you'd expect it to be. It's a stand-alone item. So if I was walking by, I might not know what it was until I saw this [info tube], and then I get it.
I wouldn't have even noticed this. I would have thought it was a safety call station.

Once people realized what an exhibit was about and they were responding to the interview questions, some people offered compliments about the experience and the aesthetics. For example:
It's cool.
It looks like a piece of art.
Made me think, 'I need to go to the Exploratorium.'
It was fun.

The poles bearing graphics and topped with water tubes were intended as icons, or "information architecture," that would help create context and continuity among the dispersed exhibits at Fort Mason. The plan was that the first signage visitors would encounter would introduce all the exhibits and establish the icon's purpose. During the summative testing, however, there were no orientation graphics, and even though data collectors briefly described the context, visitors often wondered, *What is that for?* The water tube was distracting and confusing to some people, although for others it was intriguing. As one data collector summarized it:
In the best case, everyone wanted to know what it was; in the worst case, they were misled by it, thinking that the exhibit would somehow manifest inside the water portion.

Overall, the weakest elements among all the exhibits were the graphics and text. Some people had problems seeing, reading, comprehending, and/or engaging with the labels. This was evidenced by outright criticisms of the texts (e.g., too small, too low, didn't know to turn the handle to reveal more):

The text and column could've been taller. With the sun in your eyes, it's hard to see. Makes it frustrating. You get annoyed to spend the time on it, trying to read the instructions.



More often, however, they made self-depreciating comments, blaming themselves for not understanding the exhibit or not using the signage more:

To be truthful, I didn't read it carefully.

It reminds me how bad we are at reading instructions.

It was probably user error [that prevented him from seeing the intended effect].

Keep in mind that while there were not many other common negative comments, the cued visitors represented an eager-to-please group, and the questions in the summative interview focused on probing for visitors' understanding, learning, and application of noticing skills. Feedback about how the exhibit should be improved, as would be asked in a formative evaluation study, was not encouraged.

In the section below, each exhibit will be reviewed individually for the degree to which it encouraged noticing skills or encouraged the other project goals and for any evidence of missed opportunities or problems. Because the sample sizes were small (N = 10-14) for each one, findings are reported in descriptive rather than statistical terms.

Exhibit descriptions below (in smaller text) were taken from OE planning documents. Text in italics shows a direct or slightly edited quote from a visitor or a data collector. Different quotes within one exhibit's review are from different visitors.

Architectural Mirage--As the sun heats the walls of buildings, thermal refraction causes mirage phenomena different from the more commonly seen horizontal mirages observed on roadways. Using the south facing facade of Pier 1, this exhibit provides an entry point into exploring and understanding some of the basic principles of thermal optics operating in an architectural setting.

This phenomenon was the most difficult to "get" among the seven exhibits. The mirage was hard to see if you were not sure what to look for; the instructions were complicated and took several steps; and the weather-dependent phenomenon did not always occur on cold, windy, or cloudy days. The user had to literally create the experience--often needing prompting for where to stand and what to look for. As one data collector recorded, after interviewing and observing the two visitors:
He did not see the effect. His female companion did. He read the Herbst side of the pole and then stood at the first strip. He created a line of vision between his hand and the distant visual target. He stood for more than 1 minute trying the effect. He returned to the pole to read and then went to the second foothold. Meanwhile his companion bent down to see the effect. She called out "I see it, I see it!" and then directed him on where and how to stand to see the effect.

Nevertheless, many subjects who looked at Mirage gave evidence of noticing at least part of the phenomenon. People used the terms "heat ripple," "distortion," "the ways light can bend," "refraction," "optics," "mirage," and "how your mind and eye can be tricked to see different things" to describe their experience.

The exhibit developer identified these examples of feedback as being most closely aligned with the exhibit's noticing-skills goals of tracking, correlating, valuing, and applying complex scientific principles:

I wouldn't think to look for an optical illusion, you're saying mirage, off a building. Can you do this elsewhere? [Her companion stated, "Maybe about what you can see if you really look for things."]

It is like driving on the highway; like I-5 when you look out on the horizon and see that.

It did remind me of mirrors at a carnival that create distortion.

It shows the heat reflecting off the building and how it distorts. It helps you to take more time to look and see things; be more aware of what you're seeing.

It's like when you stick a pencil in water. Light bends in different media. Air is one density and water, another. It will bend at that interface.

Visitors at this exhibit spent longer times than at other exhibits. Ordinarily, longer times would indicate more engagement, sustained interest, and exploration. In this case, the long time was usually spent struggling to see the phenomenon. One person wondered "why it was so hard to figure out." Another pondered "...if I'm just stupid and can't figure out the exhibit."

One person misunderstood the purpose of the exhibit to be to "inform people what the buildings are all about." The likelihood of a person anticipating that signage outdoors at Fort Mason might be about the history of Fort Mason (as in other national park interpretation) is to be expected. This happened at Wave Tracing and Speed of Sound as well.

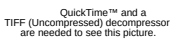
One person commented that the "kiosk" (the graphic pole) looked like it was for advertising and thought that the water on top was cloudy and unattractive. Another tried to see the phenomenon by looking through the water tube.

Speed of Sound--Using a light and a bell positioned on the end of Pier 3, this exhibit enables visitors to investigate how our perception of sound outdoors is altered by distance, wind, and temperature. Another aspect of the exhibit allows visitors to use their cell phones to time how long it takes a fog horn signal from the Golden Gate Bridge to reach different parts of the city.

This exhibit accomplished its goals with most of the visitors, but there were orientation, design, and context issues:

• People had challenges with locating the light and the bell that were 700 feet away from the graphic pole. In a few cases, visitors struggling to find them were assisted by other visitors nearby. As a data collector observed:

Two women were pushing bell button and not finding bell. Each looking at/reading different side of label, then switch sides. Then a security guard drives up and says, "You push the button, and it rings the bell over there" (pointing).



• Data collectors also saw that some visitors were directionally challenged, i.e., didn't know which way was west as instructed in the label to see the bell.

• As was heard at other exhibits, some people didn't think the exhibit was obvious enough. But another visitor liked being surprised by it:

I think it's great that you can stumble upon it.

• Four visitors focused on the water on the top of the graphic pole. In one case the data collector noted, “She instructed companion to put his ear against the water column.” One visitor was pushing the button and said, “I thought it would make it turn and maybe make bubbles.” Another seemed to entirely miss the point about the speed of sound but liked the water: "The most exciting thing was the Lucite [pillar] and the way it reflects ." Perhaps visitors focused more on the pole at this exhibit because the graphics were so far away from the other components of the exhibit.

More people at this exhibit than others said they did have some prior interest, knowledge, or training associated with the sound/distance phenomenon. Five of the 10 people interviewed gave some connection, including four with some physics education. The other person said, “I'm interested in sound design for theater, so that's interesting to me.”

People used the terms “frequency,” “weather and temperature,” and “speed of sound” when describing the exhibit and their experiences. The exhibit developer identified the following examples of feedback as being most closely aligned with the exhibit’s noticing-skills goals of precision, translation, and probing. Included were several examples of tracking--comparing this with other examples of the phenomenon they knew about:

It brings a lot of things (like the relationship between sound, light, and perception) that you hear about, but don't process.

It helps people put a name to something they’ve experienced.

Even though I knew sound travels quicker, it's not something you get to see. It's like cannon fire, as the sound takes longest.

The classic is the crack of the baseball bat when you have bad seats at the stadium.

In the service, you see the flash of light [of gunfire], then the sound, and then you feel the effect when a howitzer goes off.

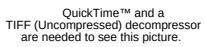
I guess it made me wonder about what today's environment, right now, would do to the speed. Is it different than another day?

This exhibit reinforced an experience that many had already noticed somewhere else, and it afforded a new experience for someone who hadn't.

The purpose of the exhibit is to help people develop better perception about sound and how weather or temperature affects that. I never really noticed that or thought about that. That's really cool. It will help me notice these things more.

No one called the bridge. One person who apparently had read the label said, "I would like to see [the effect] over a longer distance, like from across the other side of the bay." This aspect, considered one of the coolest features of the exhibit by the developers, was unfortunately lost on the visitors.

Corrosion Wedge--This exhibit illustrates the power of the spalling process by placing a piece of iron in the cleft of a specially designed block of concrete, allowing the ensuing rust to slowly fracture the block over time.



This exhibit seemed to communicate quickly, clearly, and meaningfully to visitors. People used the terms "rust," "expansion," "concrete," "crack," and "force"; they related it to things around them and were reminded of things in

their prior experience; they thought it was cool; they wondered how long it would take to break apart. In addition to simply noticing and understanding what it was about, people gave feedback that showed they were using noticing skills such as precision, valuing, motivation, tracking, probing, and challenging assumptions:

Frankly, I had no idea how strong rust is, and how much it could impact things, how rusting can be a corrosive force and how powerful it can be. I had noticed our poor buildings falling apart before but I didn't know that it was the rusting inside that was causing that.

Rusting is common; you don't stop to think about how things might expand or contract. How common things will affect everyday things. My daughter was just saying that fences have cracks because they have rusty nails.

You can apply it to home; it makes you pay more attention.

I'm thinking differently about an ordinary process and see it as a powerful influence on our environment. It'll help me to see rust as a force, that helps me understand decay as an active process.

I didn't realize that iron expanded and made cracks.

It makes me wonder if it was left forever, would it completely break apart?

Commenting on how the simple-looking wedge was actually demonstrating a powerful concept and phenomenon, people said:

It's a little subversive.

It does its job very well.

I know the Explo is kid-friendly and hands-on, but this is more cerebral. It isn't science 101A, it's science 101B! [i.e. the next level up in sophistication].

One person wanted it to stand out more: "It needs to announce itself more effectively, saying 'This is an exhibit! Come look!'"

While Wedge was a fairly passive immediate experience, the potential for active noticing from different vistas of other Fort Mason buildings, and under many other environmental conditions, was called out by visitors, although maybe this one sounds more like simple noticing:
It's to teach about rust. There's a lot of rust around here.

Wave Oscilloscope--In this exhibit a loose fender pier piling on the east pier apron of the Festival Pavilion becomes an instrument for demonstrating the complex forces converging on pier structures in the bay. A stylus attached to the oscillating piling graphs the patterns of the surrounding waves, currents, and tidal flows in a constantly changing sand tracing.

At Wave, people used the terms "wave," "forces," "pattern," "currents," "pier," and "pile." They were reminded of the Etch-a-Sketch toy and of seismographs. Unique to Wave, some people saw it as meditative.
It could almost put you in a daze to watch.

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

The exhibit designer of Wave selected the following long quote from a man (who said he was interested in physics) for the noticing skills of seeing interrelationships between physical forces and technological objects, making comparisons between two or more things, and venturing a hypothesis:

The exhibit is about seeing wave impact on this particular pier, tracing forces, preferred (wave) direction, at least for today. It's to help people understand the amount of engineering that goes into building piers, and to understand current and wave pattern: How pilings act like shock absorbers, like high-rises have shock absorbers (rollers) for earthquakes. It's obvious these things (pilings) aren't engineered equally (comparing to the earthquake absorbers), but they are absorbing energy.

Another visitor who also gave an usually long answer was an architect who also saw the interrelationships of complex systems:

It's an instrument that records motion. If you're so concentrated, you can look at it as microscopic output in these confines. But I don't think it's designed in a causal way; i.e. do this and this happens. It's designed to make people aware of their environment and relational connections.

Some wondered about what changes they'd see with different weather or an earthquake, or ventured a probing or correlating question:

I was real curious about if the piling moves up and down, ever.

What are you seeing, tide coming in or out, do you get something rebounding off those rocks?

Two others saw intrinsic value and motivation in the experience:

There's more to perceiving than you really consciously think of day to day.

I never realized how much wear the current had on it. I mean, it just brought that awareness. It also makes you stop and look around! Things you never noticed before.

It translates the motion of water into something people on land can see in fixed form.

I never realized you could do an experiment on an old pier used to prevent ships from bumping into it.

Other questions/thoughts went beyond noticing the phenomenon:
Can we use the force of waves for alternative energy?
How much longer these piers will last, and how much will it cost to maintain them?
The amount of engineering it takes to keep this landmark in good condition.

Two people had questions that revealed some uncertainties:
What is the Explo doing with this exhibit? What impact does it have?
What is the purpose of this?

No one made reference to the four diagrams of stylus patterns portrayed on the graphics or how those diagrams related to what they saw in the sand.

Bridge Thermometer--Here visitors can view the Golden Gate Bridge through a calibrated telescope that measures the height of the span as it responds to changing temperature patterns and traffic loads.

Most of the people who used this exhibit got the idea about the dynamic relationship between temperature and changes to the bridge. Noticing skills included tracking, translation, and precision:
It is trying to show the bridge expansion by temperature, to see that the bridge is flexible. I thought the bridge was flexible, but didn't know it could move due to temperature. On the anniversary (of the bridge) people walked across and it got flat, but that was because of the weight, not the temperature.
I never realized how large the effects of temperature can be on objects or buildings I take for granted.
I knew it moved side to side, but not vertical.
It shows the rise and fall according to temperature, because the cable length changes. Right? I think it's very interesting. I never thought about it until now.

People liked it and were engaged with looking at the bridge:

It's really cool.

It's a neat thing to look at.

I never realized it was that beautiful. I never realized that the rust color really stands out.

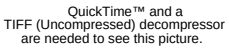
I'm amazed by the extent of change.

It's nice! I hope it's not going to be vandalized.

Several people had misconceptions or were not sure about "getting it." The viewfinder shows only the center of the bridge, and the vehicle movement is eye-catching:

Maybe it looks at warm things going by, like trucks. I don't know about the sides of the bridge.

I wasn't sure which part of the grid was supposed to match which part of the bridge.



An example of simple noticing, not related to evidence for a broadened sense of noticing or of a noticing skill, would be these comments:

It's to show how up close and personal the GGB is. To see the structure itself, to see the GGB from a different perspective... With the naked eye, you can see the beauty, but this makes it more interesting though.

The exhibit had a serious flaw. Its design invited people to move it, to use it to scan rather than to have a single point of focus. This was mystifying, distracting, and confusing to visitors:

Why does it look like you're supposed to move it, then, if you're not supposed to move it? (Friend: "So if you're not supposed to move it, why does it have finger grips?!") Maybe just put a sign that says it doesn't move.

It makes you think you could do something, but you can't, so it's frustrating.

Wind Arrows--Using sailboat wind indicators arranged at regular intervals on a vertical pole, Wind Arrows reveals the complex wind stratification patterns of the San Francisco Bay shoreline. Interpretive graphics encourage visitors to transfer their observations from *Wind Arrows* to the adjacent urban skyline. By comparing the directions of waving flags at different heights and locations along Hyde Street Pier, Fisherman's Wharf, and Aquatic Park, the surrounding urban architecture becomes a large-scale wind observatory.

Data collectors remarked that this exhibit held visitors' attention for only a short time. It had more "low" scores for time than any other exhibit. Several people missed seeing the small interpretive label. Answers were often brief. People simply noticed that it showed the variation in wind direction. Half said, "I don't know..." in their feedback along with other comments as they stretched to please the data collectors (courtesy bias).

An example of simple observation, not a noticing skill, was recognized by the visitor:

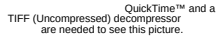
To show the direction of the wind. These are all mundane answers, huh?

Three people connected the exhibit to their interest in sailing:

If you look at any of those boats, they all have one at the top of the mast.

I'm thinking about sailing. Are the wind arrows better (at reading wind) than telltales (on sailboats)?

For all the people who do aerodynamic things, like sailing, or flying, for them it's important to know where the wind is coming from.



Even with the brief answers, the exhibit designer found evidence of noticing skills, including challenging assumptions, venturing a hypothesis and probing, and translation:

It shows the variance of the wind at different levels. I think people think the wind is always going the same way. I'm assuming it would go faster if it was really windy.

If they are 180 degrees apart, does the pole have any effect on that?

Why does the direction change every one foot?

It gives a voice to something you can't see. Or, a body to it. I can "see" wind now, thanks to the arrows.

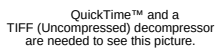
Unfortunately, the location of Wind Arrows did not allow people a skyline view with flags to notice and compare.

Audio Posts--Radio transmitters in the parking lot allow visitors to tune their car radios to hear short audio guides that deepen their experience of Fort Mason. One audio post provides a guide to common fog signals, while another helps visitors distinguish the different types of calls made by the gulls that roost on the pier sheds along the waterfront.

This exhibit posed some challenges to the data collectors because there was no physical exhibit to invite people to engage with or an associated phenomenon to give it context. Recruiting took place in the Fort Mason parking lot--an odd spot to intercept visitors--and data collectors had to convince people to get back in their cars and turn on the

radio. Refusal rates were higher for this exhibit than any other. "Only the extremely willing agreed to participate in this evaluation," said one data collector.

Four of the 10 visitors interviewed listened to only one of the two audio programs. "No one tried getting out of their car to listen for these things in real-time," said one data collector. Two people complained about having to listen from their cars and wished that there could be another way to hear the audios. Four of the 10 said they had noticed the sign at the parking lot gate.



As with other exhibits and visitors, most people were nice, eager to please the data collector, and said they thought the audios were fun, interesting, and educational.

People showed evidence for noticing skills of precision, recognizing patterns, valuing, translation, and motivation:

I've listened to foghorns all my life; never knew why I didn't notice the different sounds, that foghorns are specific to locations.

It tells me what the environment is saying to me. It's to focus on what's in the background and bring it to the foreground. We hear this all the time, and don't pay attention. I didn't know specifics of sea gull communication. I knew there were foghorns, but not different kinds.

Well, it was short, but actually a very evocative and moving few minutes.

Some visitors expected the audio programs to be about Fort Mason itself, i.e., its history, while others associated in-car audio with traffic announcements or conditions. The lack of context and visitors' expectations about the purpose were a source of confusion.

Overall Output Goals

As mentioned earlier, the characteristics and qualities that the OE exhibits should have (outputs) were fundamental appeal; interesting form and content; active participation; innovation and originality; balance of content and aesthetics; and alignment with and enhancement of the Fort Mason site. From visitor feedback and other observations of visitor behavior, some evaluative judgments about these qualities can be made. But due to the unique nature of the individual exhibits and the small sample sizes, generalizations are not easy to make for all of them. The comments below are more of a critical review by the evaluator rather than a systematic, quantitative assessment.

• Fundamental appeal and interesting form--Wave Oscilloscope, Corrosion Wedge, Bridge Thermometer, and Wind Arrows each had a physical presence that visitors could see and relate to immediately. Wave and Bridge seemed to be the most intriguing to visitors, and data collectors had few problems getting visitors to notice that they existed. For example, at Wave: "I like the way the materials blend in but still draw attention to themselves; the curved railing makes people more sensitive to what's around them." Wedge and Arrows were easier to miss. As one data collector remarked, "Although I was standing directly in front of Wind Arrows, with the sandwich board pointing directly to it,

people still asked, “Where is it?” To experience the phenomena of Mirage, Audio, and Sound, visitors had to follow instructions in the graphics on the posts or have their experience mediated by others.

• Appealing content--The appeal of the exhibits’ content was evidenced by someone saying, literally, “It’s cool” or something favorable about the content for all of the exhibits except Mirage. One person said, less directly, "I love little things; it's like bumping into things that test the mind." Whether visitors really found the exhibits and content fundamentally appealing was compromised by the cued method of the evaluation and by their own tendency toward a courtesy bias in their responses. If uncued visitors to Fort Mason had been observed using the exhibits on their own, we would have more data to answer this question (see “Limitations” above for why this was not done.)

• Active participation--Some exhibits afforded more activity than others, but all of them encouraged noticing--from making simple observations to exercising complex noticing skills. Some exhibits were easier to use and “get” than others. For Audio, active participation meant listening and having interest sustained to listen to more than one brief audio program. Arrows, as noted earlier, held visitors’ attention for shorter periods than the other exhibits. Different wind conditions on different days would increase active participation. Visitors at Mirage showed active participation for longer times, but much of it was spent being uncertain about what to do or notice. Many visitors at Bridge were also not certain about the purpose or if they were using it correctly, but there was definite motivation to try to figure it out given the intriguing nature of the exhibit unit and its focus on the famous San Francisco icon. The exhibit would be experienced best under varying weather conditions and temperature differences to notice the changes in the height of the bridge. Its location on a popular bicycle and walking path frequented by regular users made this possible. The activity afforded by Wave was quite unusual--more right-brain meditative--but it also prompted visitors to seek answers to engineering questions about the pier and pilings rather than notice immediate phenomena. At Sound visitors had to push a button, hear a bell, see a light, and notice the relationships between distance, time, and weather--requiring multiple visits under different weather conditions to discern the subtle differences. While Wedge

was a more passive immediate experience, the potential for active noticing at other Fort Mason buildings as well as other public buildings, bridges, fences, homes, mining equipment, and ships was noted by visitors.

Overall, the exhibits offered few opportunities to actively and concretely manipulate variables on the spot, observe new results, and draw conclusions. Instead, the noticing skills of transferring, venturing a hypothesis, gaining precision, and challenging assumptions were encouraged.

• Innovation and originality--Many of the comments by visitors suggested that they were struck by the originality of repurposing objects and activities as exhibits at Fort Mason: the yacht telltales on Wind Arrows, or an old piling to direct a stylus in sand patterns for Wave, a more precise view of the Golden Gate Bridge, and rusting-in-process for Wedge. The innovativeness was sometimes obscured by unclear directions in the texts or affordances of the exhibit (e.g., for Bridge, Mirage, Sound).

• Balance of content and aesthetics--This was difficult to assess, but in one case, with Wave, there seemed to be a disconnect between the affective enjoyment for visitors provided by the abstract, mesmerizing moving stylus in the sand and the rigor of the patterns predicted in the graphics, which were not usually visible in the visitors' direct experience. The audio of the foghorns provided an unexpectedly powerful balance with the evocative auditory experience that was both beautiful and informative.

• Alignment with and enhancement of the Fort Mason site--One of the requirements of the Exploratorium's use of Fort Mason was to create exhibits that would enhance visitor experiences outdoors but not call attention to themselves or compete with the ambiance of the park's historic setting. Clearly, from visitors' comments about the lack of obtrusiveness of the exhibits--to the point of being almost invisible, especially without any orientation or context to the fact that they were installed there--this intent was successfully accomplished. Even though visitors were told that these

were exhibits made by the Exploratorium, some people stuck to their assumptions and expectations that exhibits outdoors at Fort Mason would be like other national park “wayside” exhibits that do not have a purpose beyond interpreting the immediate natural and human history of the site (e.g., original purposes of the buildings, shipping, war and military).

DISCUSSION AND RECOMMENDATIONS

This section provides a preview of what issues and possibilities exist for another version of outdoor exhibits when the Exploratorium moves to a location near Fisherman's Wharf. It also informs the remaining exhibits to be installed as well as the evaluation of the whole collection in the second summative evaluation. Three topics will be briefly discussed: the utility of noticing skills, the importance of context, and the indispensability of formative evaluation.

Noticing Skills--All seven of the exhibits evaluated in this study succeeded to one degree or another at encouraging noticing and promoting noticing skills with visitors. The exhibits could easily be relocated to another site where other people could be engaged (uncued), providing the phenomena are present, e.g., a loose piling, a nautical bell or horn marker, a south-facing building, a view of the Golden Gate Bridge. Better text and graphics (e.g., larger type size, clearer instructions, text not wrapped around a pole) would make exhibits easier to use and understand, and eliminate some visitor frustrations.

Among the intended goals, noticing skills were the strongest outcome with the participants in this study. Enabling noticing skills was an unusual and exciting experience for many people. This goal is very suited to helping visitors feel competent and interested in outdoor natural phenomena--a worthy visitor outcome for many science museums. The OE exhibits can serve as good models for what is possible.

There was less evidence for the other two goals of helping visitors explore "complex systems and interactions at play in an outdoor environment" and "come to a deeper understanding of the phenomena by applying scientific concepts and principles to the outdoor environment"--and this was with cued, willing visitors who spent a couple of minutes engaged with the exhibit. Better orientation to the collection as well as exhibits located in a less-dispersed manner would help give visitors more related examples to explore and build a more complex scientific generalization upon. Those two goals, however, are very lofty.

The methods and evaluation process for this subset of OE exhibits has helped clarify and define what was meant by "noticing skills" and can be further refined in the next round of summative evaluations.

Recommendation: Keep noticing skills as a primary outcome for the next generation of OE exhibits. If the phenomenon to be noticed will not always be present, due to certain weather conditions such as low temperature, no wind, or the presense of fog, make that fact clear to visitors in the labels.

Context--If the exhibits had been installed outside the Exploratorium's home at the Palace of Fine Arts instead of at Fort Mason, they would have benefitted from a more logical context. Visitors to Fort Mason not unreasonably expected the exhibits to be more concretely related to *that* site. Moreover, many of the visitors who came to Fort Mason were there for reasons unrelated to seeing exhibits outdoors. They were on their way to or from an event, and that's where their attention was focused.

The summative evaluation used feedback from visitors who had experienced only one of the seven widely disbursed OE exhibits. In a context where there were multiple exhibits closer to each other and visitors could easily move from one to another, visitors might have made more connections between phenomena and been able to apply the scientific concepts they had in common. There, visitors would assume that the exhibits were made by the Exploratorium and were related to the Exploratorium's themes of science, art, and human perception. As one OE team member commented, however, "I think the clustering can help, but I'm not at all certain about helping people see shared principles." The evaluator agrees.

Recommendation: Review the summative evaluation results in light of the challenges that the lack of context caused, and look forward to overcoming them in the new installations. Perhaps scale back on the goal of visitors understanding complex scientific principles and stress that the exhibits scaffold more on people's prior knowledge.

Formative Evaluation--Few interactive exhibits make it to the museum floor in a form that does not need and greatly benefit from testing with the public. What is obvious to exhibit designers can be obscure to exhibit users. Even the briefest tryouts can reveal weak points in design or interpretation. Issues of appeal, apprehendability, and affordances can be improved and refined to benefit visitor experiences and outcomes and to increase the likelihood of achieving the exhibit's goals. Summative evaluations of exhibits that have not been through any formative design improvements will reveal missed opportunities and misinterpretations. Those challenges can make visitors feel incompetent--and visitors should not blame themselves for poorly written exhibit instructions.

With outdoor exhibits, the issue of durability is compounded by sun, heat, humidity, rain, insects, trucks, etc., which can be added to the list of possible sources of trouble and abuse. Many of the OE exhibits in the first subset group had not been installed for long, and breakdowns delayed the data collection more than once. The challenging outdoor environment was very different from what the Exploratorium staff was accustomed to.

The original grant proposal called for extensive iterations of formative evaluation with the OE exhibits. Front-end and formative studies were conducted on some earlier prototypes at other sites, and these gave useful information that could have informed exhibit development at Fort Mason. For example, formative tests at Rincon Park indicated that some visitors thought that the exhibits were broken when they couldn't observe a phenomenon that was simply not present (wind) at the time. These tests also revealed a very different audience from the people who visit the Exploratorium at the Palace of Fine Arts--a higher refusal rate, for one thing. (Separate studies are available from the Visitor Research and Evaluation Department at the Exploratorium.)

Working at Fort Mason--when that site was finally selected--involved two partners, the Golden Gate National Recreation Area and the Fort Mason Center, and the OE exhibits needed approval from both of them. The GGNRA

approval process alone involved review from five departments, each with their own system (Cultural Resources, Natural Resources, Project Review, Signage Review, and Accessibility). With a compressed time frame due to site selection issues, high turnover on the OE team, and the additional approval process, the project fell short of the grant's original formative evaluation goals. By the time of this pre-summative report, some form of testing had been done on five of the seven exhibits, and the information obtained was being used to develop the remaining part of the collection to be installed at Fort Mason.

Recommendation: During formative testing, 80% of the visitors should be able to use, understand, appreciate, and have appropriate outcomes with the OE prototype. The final version of the exhibit should maintain the successful qualities discovered and refined during formative testing.

The findings from the seven exhibits evaluated at Fort Mason can inform the planning, development, and evaluation of new and relocated installations for the Exploratorium or for other museums or science centers that want to construct outdoor exhibits that feature natural phenomena.

THANKS to all the visitors who participated in the interviews at Fort Mason. Thanks to Leah Johnson, Nancy Carlisle, and Fay Dearborn for collecting and transcribing the data; to Toni Dancu for asking all the good questions about the evaluation study; to Adam Klinger for advising and coordinating the evaluation training and schedule; to Bryan Connell, Shawn Lani, Charles Sowers, Maz Kattuah, Ulrika Andersson, and Steve Gennrich for the exhibit design, installation, and feedback on the evaluation plans, data, and reports.

APPENDIX FILES (ELECTRONIC)

OE METHODS--Data sheet, recruiting statements.

GOOSE BUMPS--Transcriptions of interview feedback for each exhibit with highlighted “goose bumps” (statements that resonated most with the exhibit goals) by exhibit developers.

MASSED OE SUMMATIVE DATA--Individual Excel files with original data for each exhibit grouped by exhibit.

FORMATIVE OE DATA--Transcriptions of data from data collectors for each exhibit. Not massed, and not reviewed in this report.