Children, Youth and Environments 21(2), 2011

Using Iterative Design and Evaluation to Develop Playful Learning Experiences

Toni Dancu
Joshua P. Gutwill
Nina Hido
Exploratorium
San Francisco, California

Citation: Dancu, Toni, Joshua P. Gutwill and Nina Hido (2011). "Using Iterative Design and Evaluation to Develop Playful Learning Experiences." *Children, Youth and Environments* 21(2): 338-359. Retrieved [date] from http://www.colorado.edu/journals/cye.

Abstract

This paper explores the importance of *iterative design and evaluation* in developing playful learning experiences in museums. According to research, play has five defining aspects: it is structured by constraints, active without being stressful, focused on process not outcome, self-directed, and imaginative (Gray 2008). For each of these aspects, we demonstrate how an iterative process of development and formative testing improved several museum exhibits, engendering more playful learning experiences for visitors. We focus on the assessment element of the design-test process, offering in detail two methods for gaining user feedback from visitors: interviews and observations. Interviews capture visitors' self-reports of their playful experiences, while observations record actions and conversations during play. Examples illustrate how these methods may yield information useful to development teams when revising educational experiences. The paper concludes by considering some of the challenges of enacting design-test processes and by offering potential solutions.

Keywords: iterative design, formative evaluation, informal education, museum, spatial reasoning, exhibit, science, math, play

Introduction

In free-choice educational environments like museums, learners will engage with activities, exhibits or programs only as long as those offerings hold their interest. Playful experiences—those that are pleasurable, self-directed, and open-ended—are therefore powerful motivators for sustained exploration and learning to occur. This paper explores the importance of *iterative design and evaluation* in developing playful museum experiences that effectively promote learning. The development process will be described in the context of *Geometry Playground*, a traveling museum exhibition created by the Exploratorium, San Francisco's museum of science, art and human perception. *Geometry Playground* encourages visitors to practice the cognitive skill of spatial reasoning by playing with geometrical phenomena in an array of interactive exhibits.[1]

Play-based experiences that simultaneously spark visitors' interest while engaging them in learning may benefit from folding visitor feedback into the design process (Taylor 1991). At the Exploratorium, visitor researchers and evaluators typically conduct interviews, real-time observations, and video observations to inform exhibit development. During formative evaluation of *Geometry Playground* exhibits, we assessed the ways in which visitors were playing and reasoning spatially. That information was used by development teams to improve final exhibit designs (Patton 2008). A concrete example may help illuminate the process. One of the *Geometry Playground* exhibits entailed a large cylindrical mirror and a hopscotch graphic mathematically stretched across the floor so that the hopscotch grid's image appeared conventional in the cylindrical mirror. This exhibit underwent multiple iterations informed by formative evaluation. The exhibit began as a four-foot high cylindrical mirror at whose base lay a large distorted photograph of a kitten (see Figure 1). In the mirror, the kitten morphed into a surprisingly perfect image, leading visitors to remark on its cute but astonishing appearance. Conversation about shapes and space would have been considered evidence of spatial perception and reasoning (Leinhardt and Crowley 1998); however, formative evaluation indicated that spatial talk was largely absent from the kitten version of the exhibit. The team experimented with several potential graphics, including different objects, words, and activities. Eventually, a hopscotch activity was chosen based on team preferences (see Figure 2). According to formative interview studies, the hopscotch activity increased conversations about the shape and transformation of the floor graphic. Real-time observations found visitors returning to the exhibit and calling others over to experience it, which the team took as evidence of visitors' intrinsic motivation and interest in the process (playing hopscotch) over the outcome (successfully completing hopscotch). In summary, the exhibit began as an experience of viewing a surprising but static image. By engaging in a design-test cycle, the development team transformed it into a socially playful activity that promoted talking and thinking about geometry.

---

[1] In our nomenclature, an *exhibition* is a collection of *exhibits*, which in turn offer museum visitors activities or phenomena to explore. An *exhibit element* is a piece of an exhibit, usually one that cannot stand on its own to engage visitors.

Figure 1. Initial kitten graphic (photo courtesy of the Exploratorium)

Figure 2. Final hopscotch graphic (photo courtesy of the Exploratorium)

The remainder of the article begins by describing one type of museum learning fostered in *Geometry Playground*—practicing the skill of reasoning spatially—followed by an introduction to the important aspects of play and how these aspects are reflected in the museum field. We go on to present our design-test cycle and two methods we often employ as part of that cycle. We provide examples illustrating how these methods assessed each aspect of play and informed the development of specific exhibits. Finally, we consider some of the challenges of enacting design-test processes and offer potential solutions.

Promoting Playful Learning

The Exploratorium recently completed *Geometry Playground* to help visitors practice spatial reasoning skills by enticing them to explore geometric shapes and play in unusual spaces. Spatial reasoning in this context involved delighting visitors in "understanding, manipulating, reorganizing, or interpreting relationships visually" (Tarte 1990). We hoped to foster Linn and Petersen's (1985) three main facets of spatial reasoning: (1) spatial perception, which involves identifying spatial relations relative to oneself (used in hide-and-seek or climbing play structures); (2) spatial visualization, which entails manipulating complex spatial information over several steps (needed in Etch-a-Sketch or other drawing games); and (3) mental rotation of two-dimensional and three-dimensional figures (utilized when solving jigsaw puzzles or Rubik's cubes). Practicing thinking skills like spatial reasoning is an important part of science learning, albeit different from understanding concepts (National Research Council 1996; 2009). In addition to promoting playful spatial reasoning, the goals for the exhibition included engaging and enchanting visitors with geometrical phenomena, goals that align well with various aspects of play. After all, museums should be designed to be not only educational, but fun (Packer 2006).

Many theories and definitions of play share common features (see Hutt 1981; Meisner 2007; Piaget 1953; Vygotsky 1978). For example, the *Dictionary of Education* defines *play* as "any pleasurable activity carried on for its own sake without reference to ulterior purpose or future satisfactions" and the *children's play* entry adds, "freedom of action among children, for pleasure and amusement" (Good 1973, 426). The most recent *Handbook of Child Psychology* purports, "the essential benefits of play lie in the manipulation of information in a pressure-free context that is informed by external and internal determinants, but not controlled by either" (Rathunde and Csikszentmihalyi 2006, 508). Many components of these definitions are reiterated by Peter Gray (2008) whose five aspects of play may be summarized as follows:

1. Play has structure;
2. Players' minds are active but not stressed;
3. Play is focused on the process, not the outcome;
4. Play is chosen and directed by the player(s); and
5. Play is imaginative and removed from real-life consequences.

Gray's aspects of play fit well with our own views about visitors' interactions in the museum and provide a useful framework for illustrating in this article our iterative design and evaluation process. To create and understand playful learning experiences at exhibits in a manner resonant with Gray's key aspects, the *Geometry Playground* team employed an iterative design-test process. Next we briefly describe this process and the evaluation methods employed within it.

Evaluation in the Design-Test Process

In our development process, exhibit and label developers create a draft version of an exhibit, called a *prototype*, then evaluators systematically assess its effect on visitors' experiences. The results of the evaluation study inform the next iteration of the prototype. A key feature of the process, similar to the practice of design-based research, is the repeated cycle of exhibit development and revision (Cobb et al. 2003; Medlock et al. 2005). The overarching goal is to create intuitive, interesting, understandable, and enjoyable exhibits for museum visitors.

The evaluative part of the process requires several steps. First, the evaluator discusses the exhibit prototype in depth with the exhibit and label developers, together articulating experiential goals for visitors and raising any specific questions for visitor testing. Next, the evaluator translates those broad questions into empirically answerable ones. For example, suppose anecdotal observations suggest that only children are climbing a large-scale exhibit designed for parents and children to use together. The team decides to create various versions of a label to urge adults to join their children on the climber. Empirical questions may include: What percentage of adults near the exhibit actually climb on it? How does that percentage change when each version of the label is added? What effect does the label's placement and orientation have on adult engagement? What are some of the reasons adults give for resisting playful climbing, and can those reasons be related to the design of the exhibit?

While the questions in the above example focus on issues specific to the exhibit under investigation, evaluation studies also pursue commonly-held team questions, such as:

- Are visitors interested in the exhibit concept?
- Can visitors quickly figure out how to interact with the exhibit to produce interesting effects?
- Are visitors motivated to continue (i.e., focusing on process over completing something specific)?
- Does the exhibit spark any new areas of inquiry for the visitor?
- Are visitors practicing the skills we've designed for?
- Are visitors frustrated or confused?

In the third step of the assessment process, the evaluator chooses a method for studying the prototype, and in some cases, designs an experiment to compare different versions of particular exhibit elements. Many methods could effectively explore the general questions raised above, but those most often employed at the Exploratorium are interviews, real-time observations, and video observations. Nearly all studies of exhibit prototypes are implemented in the natural context of the museum floor to ensure strong ecological validity. Museum visitors are selected to participate via random or purposive sampling procedures, depending on the goals of the study (for a discussion of these sampling approaches in museum studies, see Allen et al. 2007). Visitors are also either "cued" or "uncued" as to their participation in the study when interacting with an exhibit. In a cued study, visitors are invited to participate *before* using the exhibit. Cued studies represent "best-case" scenarios, because visitors often act their best knowing they will be assessed in some way. Visitors in an uncued study are recruited for interview *after* they have used the exhibit, or are simply observed using it without explicit recruitment. In either type of study, evaluators explain that visitors' feedback will help the museum improve the exhibit, in an attempt to mitigate the natural pleasing bias and elicit more honest responses from visitors. To obtain useful and timely results, formative evaluation studies in a design-test cycle typically focus on problems and solutions rather than successes, include small sample sizes (10-15 visitors or visitor groups), analyze data for qualitative themes or simple descriptive statistics, and yield brief reports.

After completion of the study, evaluators and developers meet to discuss results and consider emergent problems. The development team creatively finds solutions, implements changes and starts a new cycle. Evaluation theorists refer to this collaborative effort to create evaluations with immediate utility as *utilization-focused evaluation* (Patton 2008). We have found that this iterative design-test process consistently yields exhibits that visitors find easy to use, engaging and comprehensible.

Below we describe in more detail the methods *Geometry Playground* evaluators employed—interviews, real-time observations, and video observations—along with the benefits and challenges of each.

Interviews

Exhibit interviews collect information about thoughts and feelings directly from the visitors themselves, in their own words. Evaluators may conduct structured interviews, asking participants the same set of questions in the same order, or semi-structured interviews, varying questions to probe a set of broad issues (Vissner et al. 2000). Analysis depends on the format of the questions: open or closed. Open-ended questions allow visitors to respond in their own words; evaluators look for themes and may code the responses, yielding simple descriptive statistics. Closed-ended or rating scale questions require visitors to choose from a set of options; evaluators calculate percentages, averages and frequencies, which may be subjected to statistical analysis with large enough samples (Rosenthal and Rosnow 2008). Challenges to conducting effective interviews include formulating unambiguous questions, sequencing questions to suppress researcher bias, identifying areas to probe when responses are unclear or complex, ensuring that visitors do not feel tested by questions and probes, and carefully interpreting and analyzing responses (Rosenthal and Rosnow 2008; Vissner et al. 2000).

G*eometry Playground* evaluators developed structured interview protocols to explore exhibit usability and assess goals specific to each exhibit. These protocols ranged in size, typically including five to ten questions, and employed both open and closed-ended questions.

Observations

Observations allow evaluators to assess the behaviors—actions and conversations—of visitors as they interact with exhibits. Observations may be collected in real-time, with evaluators systematically noting visitors' behaviors, or recorded using audio/video capture for later analysis. The former requires coding or categorizing behaviors during data collection, while the latter allows for coding after data have been collected. In either case, data collectors and coders analyze visitors' behaviors based on clearly defined coding schemes or observational checklists, which usually identify mutually exclusive codes (Bakeman 2000; Bakeman and Gottman 1997). In this way, data are transformed from qualitative actions or utterances into quantitative frequencies, durations or even patterns of code sequences. Depending on the complexity of the behaviors and conversations under observation, coding schemes may take considerable time to develop and refine. Specifically, simple schemes may count actions such as manipulating an exhibit element or may record time spent at an exhibit; more complex schemes may assess the degree to which visitors successfully complete an activity at an exhibit; and the most complex schemes code visitors' conversations to infer their thought processes or emotional responses. Typically, video observations afford more complex coding schemes than real-time studies, due to one's ability to rewind and focus on subtle or complicated behaviors.

In general, observations have the advantage of being less obtrusive to visitors' experiences than interviews, producing a more ecologically valid understanding of how visitors use and interpret an exhibit. One challenge for conducting observations in busy or noisy environments like science museums is that access to visitors' conversations and actions may be obstructed or compromised. Common concerns specific to video observations include the high potential for visitors' reactivity to recording devices (Smith et al. 1975), and the time-intensive process of analysis (Derry et al. 2010).

When conducting real-time observations of visitors engaged with *Geometry Playground* exhibits, data collectors positioned themselves near the exhibit. They employed an observation checklist created by the evaluator to address the specific goals of the exhibit. Checklists typically required data collectors to track the occurrence or non-occurrence of visitors' behaviors at the exhibit, as well as record overall interpretations of visitors' frustration and affect.

For video observations of visitors at *Geometry Playground* exhibits, we roped off an area of the museum and placed several signs around the area to inform visitors of the study and audio/video recording (Gutwill 2003). This set-up allowed visitors to decide whether or not to be videotaped. Cameras captured the behaviors, and often conversations, of visitors interacting with the exhibits. The video was reviewed and coded later according to schemes developed to address particular goals of each exhibit. Video data collection also allowed the development team to view the interactions again and again, and pose questions to the data set that emerged from those viewings.

We have described our design-test process for creating exhibits, focusing on methods for interviewing and observing visitors in museums. In the next section of this article, we employ Gray's theory of play to frame several vignettes from *Geometry Playground* illustrating the efficacy of the design-test process.

Evaluating Multiple Aspects of Play

Peter Gray's five defining aspects of play echo several authors' definitions and descriptions of play (Mornighan-Nourot et al. 1987), and align well with our conceptualization of the museum experience (c.f. Oppenheimer 1972). Below we describe Gray's facets of play along with additional research on play and museum learning. Gray's aspects are not mutually exclusive, but each emphasizes an important part of play. Following each aspect, we present examples of interviews and observations used to inform design iterations for *Geometry Playground* exhibits.

1. Play Has Structure, or Rules, Generated or Agreed upon by the Players

"Play draws and fascinates the player precisely because it is structured by rules that the player herself or himself has invented or accepted" (Gray 2008, 5). Even when no external rules are imposed on play, implicit rules give direction and meaning (Mornighan-Nourot et al. 1987; Sapora and Mitchel 1961). Implicit constraints exist in all museum exhibits because exhibits focus visitors' attention on particular phenomena. Oftentimes, these constraints will come in the form of a challenge, which helps motivate visitors by giving them something to work toward (Perry 1993). In the case of *Geometry Playground*, geometric rules provided such constraints. That is, many exhibits were designed to display and give visitors opportunities to manipulate aspects of a particular geometric rule. For example, stacking a set of blocks in one exhibit entailed following the constraint of tessellation—stable stacks packed tightly with no space between the blocks. This constraint of packing structured a playful experience of spatial reasoning in the exhibit.

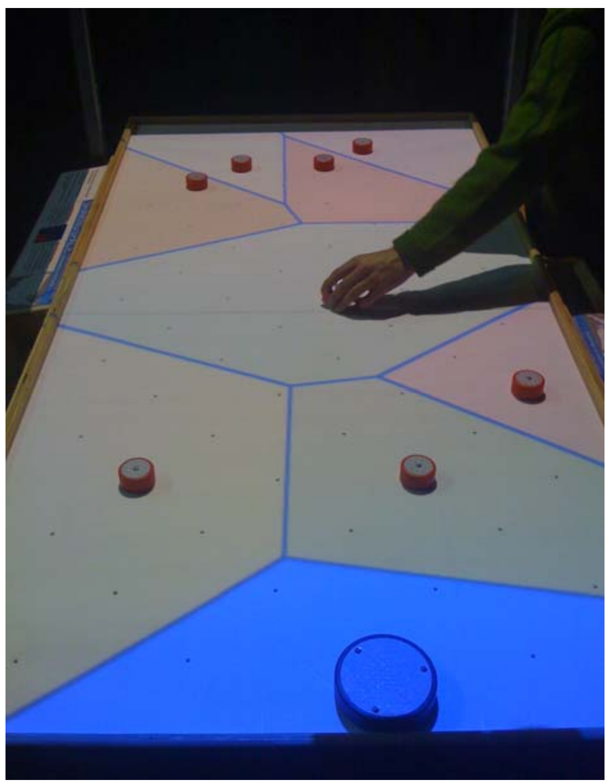
Interview

During the design-test cycles of exhibit development, *Geometry Playground* evaluators often attempted to assess spatial reasoning, a skill practiced in the minds of learners. Interviews explored how visitors thought about the spatial relationships that were implicitly driving exhibit experiences. For example, at the exhibit *Dividing Space,* visitors rearrange red pegs and roll blue pucks through their peg configurations (see Figure 3). The exhibit continually redraws the boundaries among puck and pegs according to an underlying geometric constraint—each boundary line lies equidistant between the two closest points—producing a pattern called a Voronoi diagram. This constraint drives the exhibit's phenomenon and yields the interesting real-world applications mentioned in the label (such as determining where best to locate fire stations within a city). Unfortunately, anecdotal observations suggested that the exhibit afforded off-task, air hockey game-play, potentially overpowering the learning experience at the exhibit. We employed interviews to determine whether visitors were off-task or engaging in playful spatial reasoning. We covered the explanatory labels at the exhibit and asked visitors what they were trying to do at the exhibit and what happened when they moved the blue puck. Visitors described the exhibit experience in their own terms, many talking about space, shapes, patterns, and interrelationships, providing evidence that the exhibit was indeed stimulating spatial thinking:

> *Adult male: [I was trying to] create a path and in doing so, see all the different shapes that develop.*
>
> *Child male: [I was] just looking at the polygon shapes. It was neat watching the lead circle [puck], and it created the blue [area] and changed the others as it passed.*
>
> *Adult Female: [When I moved the puck] the shapes changed especially when it hit the red ones [pegs]. The blue [area] is bigger or smaller.*

This example demonstrates the power of interviews to capture thinking that observations may miss. Despite the evidence of spatial reasoning driven by the exhibit's constraints, the development team added a label activity to focus visitors' attention on the changing shapes and areas, further increasing the spatial talk at the exhibit.

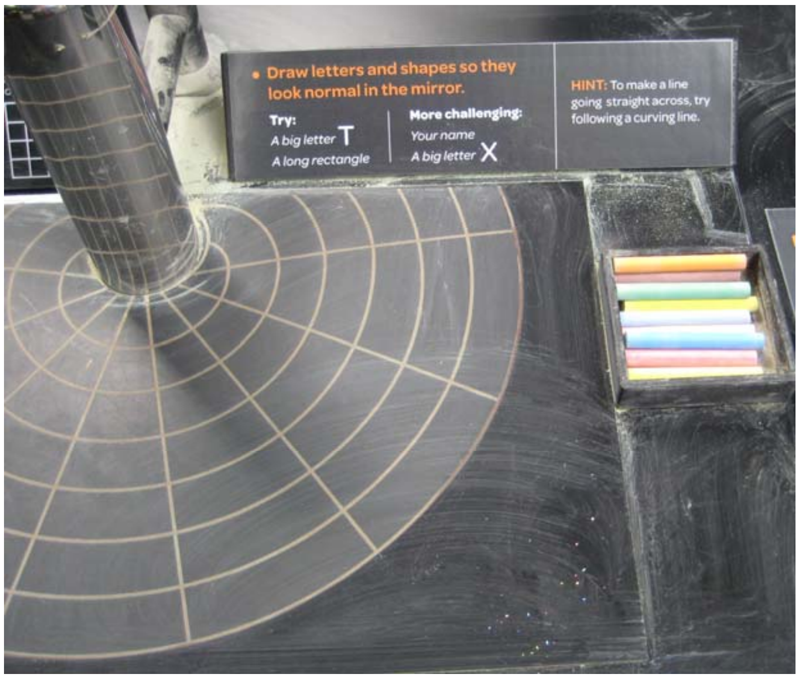
Figure 3. *Dividing Space* exhibit (Photo courtesy of the Exploratorium)



Observation

Real-time observations may also determine how implicit rules structure the exhibit activity for visitors. At the exhibit *Distorted Drawing*, visitors draw on a chalkboard while viewing their drawing in a cylindrical mirror (see Figure 4). To produce a shape that appears normal in the mirror, visitors must draw a stretched shape on the chalkboard. In a process called anamorphosis, the cylindrical mirror compresses any drawing horizontally, so straight lines appear curved and curved lines appear straight. The initial exhibit prototype included a mathematically correct curved grid painted on the chalkboard as a guide for this constraint. Over time, the painted grid faded. Systematic observations of visitors' drawings revealed that the grid was necessary to support visitors' exploration of the anamorphic rule. Based on these evaluation results, the curved grid was permanently etched into the chalkboard, providing a guide subtle enough to avoid overpowering the joy of discovery.

Figure 4. *Distorted Drawing* exhibit (Photo courtesy of the Exploratorium)



2. Players Are Active and Alert, but Have a Non-Stressed Frame of Mind
"Players do not just passively absorb information from the environment, or reflexively respond to stimuli, or behave automatically in accordance with habit" (Gray 2008, 7). Leading researchers agree that play entails dynamic, attention-absorbing, and persistent activity (Sapora and Mitchel 1961). Similarly, science museums often design exhibits to engage visitors in actively seeking answers or practicing scientific-thinking skills (Humphrey and Gutwill 2005; Semper 1990). Play researchers also emphasize that goals in playful activity typically are achievable and do not engender stress (Gray 2008; Sapora and Mitchel 1961). The iterative design-test cycle aims to identify and reduce frustration (or stress), oftentimes through strong affordances or good human-factors design (c.f. Norman 1988).

Interview
Interviews can reveal specific problems for visitors and discern whether visitors perceive frustration as an engaging challenge or a stressful road block. For *Geometry Playground*, formative evaluation interviews typically included the question, "Was anything frustrating or confusing about using this exhibit?" For

example, at the exhibit *Space-Filling Blocks* mentioned above, visitors stack star-shaped 3-D blocks so that they fit together with no wasted space (see Figure 5). To stack these shapes, visitors must carefully observe how their points will fit together and consider their spatial orientation. Interview responses initially revealed that stacking these blocks looked simple, but could be quite difficult and frustrating:

> *Adult Male: Looking at it [the block], it looks like it fits intuitively, but then it was hard, and I didn't know how to actually make them fit.*

> *Child Female: No [nothing frustrating or confusing]… well, actually the little stars [blocks] were kind of confusing. We didn't really get how they fit in the pattern.*

This feedback led the development team to reduce potential stress by including a hint for visitors in the label: “It’s harder than it looks.” The next round of interviews revealed that visitors found the exhibit difficult, but in a good, challenging way that kept them thinking, but not stressed:

> *Child Male: They [the blocks] had different sides. It was challenging, not frustrating.*

> *Child Male: It was fun… challenging, how I had to try to see how they fit, like ‘Oh this doesn't work, but this does.’*

The design-test cycle produced a minor change that helped align visitors’ expectations to the demands of the activity, reducing stress and amplifying the exhibit’s playful challenge.

Figure 5. *Space-Filling Blocks* exhibit (Photo courtesy of the Exploratorium)

Observation

Observations may assess frustration quickly by comparing the actual use of an exhibit to its intended use. If visitors cannot use or make sense of an exhibit easily, they will typically lose motivation and quit the exhibit. At the *Geometry in Motion* exhibit shown in Figure 6, visitors build their own mechanical linkages out of rods and pivots, transforming one kind of motion (e.g., back-and-forth) into another (e.g., circular). The initial exhibit prototype contained five different linkage parts for visitors to use. Data collectors observed and noted how visitors approached the exhibit, used the label, and attempted to create successful linkages. Qualitative analysis of these observations led to the following conclusions:

- There seemed to be too many different parts for visitors to use, making it difficult for them to get started.
- Visitors built linkages with little regard for which parts they used.

Over the course of several design-test cycles, the team added labels, incorporated instructional photos, color-coded the rods, and reduced the number of varying pieces. However, observations revealed that there were still too many distinct parts, resulting in the same visitor confusion. The final design for *Geometry in Motion* included two simplified universal pieces that successfully promoted strong initial engagement, motivating visitors to pursue the suggested activities or free-build for prolonged periods. Simple, intuitive exhibit design, often informed by systematic observations, can reduce stress and enhance engagement by allowing visitors to focus on the activity rather than the workings of the exhibit elements (Humphrey and Gutwill 2005; Sauber 1994).

Figures 6a and 6b. *Geometry in Motion* exhibit and close-up on components and use (photos courtesy of the Exploratorium)

3. Playful Activity Values Means over Ends

Sapora and Mitchel (1961) claim that the "essential characteristic of play is a satisfaction in the activity itself" (118). Players are concerned less with the goal as with the ways of reaching that goal (Mornighan-Nourot et al. 1987). Gray (2008) goes on to explain that play often has goals, but those goals are only part of the game, not the reason for participating in it:

> *For example, constructive play (the playful building of something) is always directed toward the goal of creating the object that the player has in mind. But notice that the primary objective in such play is the creation of the object, not the having of the object* (3).

The Exploratorium frequently creates exhibits that are open-ended, encouraging exploration rather than the pursuit of a correct answer or obvious conclusion. This type of design imparts visitors with a sense of confidence that makes them more motivated to approach and continue using an exhibit (Perry 1993).

Interview

Interviews may explore visitors' beliefs about the nature of their endeavors: are they working toward a specific goal or more playfully engaged in exploration for its own sake? In an early iteration of *Space-Filling Blocks* (where visitors stack together star-shaped blocks), images of completed block towers on the label and a limited number of blocks inadvertently communicated a specific end goal. When we asked visitors, "What made you feel like you were done here?" several reported that they felt completion after using all of the blocks or building a stack to look like the one in the image.

> *Child Male: Those two blocks didn't work [like the blocks in the image], so I stopped.*

> *Child Male: There's two extra blocks, I couldn't make a pyramid.*

Concerned that a clear end goal would shut down playful exploration, the exhibit development team revised the label to depict stacking in progress rather than a finished design, and redesigned the exhibit to support multiple ways of stacking the blocks. With these changes, the emphasis shifted from attaining a single end state to free-building in numerous orientations.

Observation

Observations may confirm when playful experimentation, without clear end goals, has been successfully encouraged. For the *Geometry in Motion* exhibit (where visitors create linkages), the developer wanted to provide a sample linkage to help visitors get started, but did not wish to interfere with playful investigation and free-building. After recording video observations of visitors at the exhibit, the evaluator and developer sat together to review visitors' behaviors. Pausing after each visitor group's interaction with the prototype, the two discussed whether visitors were free-building or simply reconstructing the sample linkage before leaving. Unfortunately, several groups enacted the latter behavior, so the design team tried including three different sample linkages along with a suggestion that the samples were meant to "get you started." Visitors began building their own linkages. Using video observations, the design-test process identified and removed obvious end goals, resulting in a more varied, free-building experience at the exhibit.

4. Play Is Self-Chosen and Self-Directed, Continuing Only as Long as the Player Wishes

"Players not only choose to play or not play, but they also direct their own actions during play…The ultimate freedom in play is the freedom to quit" (Gray 2008, 2). This genuine and continued interest in an activity is often referred to as intrinsic motivation (Csikszentmihalyi and Hermanson 1995; Mornighan-Nourot et al. 1987; Packer 2006). In a review of the research on play, Sapora and Mitchel (1961) purport that play has educational potential only when the player finds it interesting and exciting. Similarly, learning in museums is self-driven, based on visitors' interests and motivations. Indeed, researchers often apply the term "free-choice learning" to museum experiences to emphasize a visitor's freedom to choose to participate, act as one desires, and quit when finished (Falk and Dierking 2000; Perry 1993). While visitors will always exercise choice and pursue their interests in a museum, some exhibits may support open-ended, self-directed play better than others.

Interview

Interviews may measure the extent to which visitors feel self-directed and free to choose or leave exhibits, though observations are more typically employed. In the case of *Geometry Playground*, our formative evaluation team did not find a need to use interviews to assess self-directed engagement. However, previous Exploratorium projects have used interviews to address this issue. For example, in the *Active Prolonged Engagement* (APE) project, interviews conducted as visitors left exhibits asked, "We're interested in finding out what makes visitors move on from one exhibit to another. Thinking back on it, what was it that prompted you to move on to the next exhibit?" This question was posed to gain an understanding of whether visitors were leaving an exhibit for extrinsic factors (those not related to the exhibit, such as needing to break for lunch) or for intrinsic reasons (those related to the exhibit, such as having used all of the exhibit elements). APE exhibits were designed to be open-ended, providing opportunities for engagement as long as visitors chose to continue playing. Results showed that significantly more visitors gave extrinsic reasons for leaving APE exhibits than for exiting classic Exploratorium exhibits (Gutwill 2005).

Observation

Video observations may yield highly accurate data about visitors' behaviors, useful when pursuing questions of self-direction in play. Frequently, museums evaluate visitors' engagement by means of two unobtrusive observations: attraction, whether or not the visitor stops at (chooses) an exhibit; and holding time, the length of time visitors spend (continue to play) at an exhibit (Falk 1983; Serrell 1998).This was the case for *Geometry in Motion* (where visitors create linkages). The initial prototype anecdotally appeared more engaging to males than females. To check the validity of this perception, we employed video observations to measure comprehensively the percentage of males and females walking near the exhibit who actually engaged with it. Real-time observations would have been difficult, due to multiple exit points, high crowding and return visits to the exhibit area. Analysis of the video data found a significant gender gap, with more boys using the exhibit and spending more time at it. Based on prior research and colleagues' wisdom, the developer team altered the exhibit to better attract and engage girls. Another round of video observations found that significantly more girls chose to stop at the "female-friendly" version of *Geometry in Motion* and spent more time with it, while boys continued their high levels of engagement. The iterative process of development and testing produced an exhibit that better fit girls' interests and choices without harming boys' experiences.

5. Play Is Imaginative, Non-Literal, and Mentally Removed from Serious Life

"In play one enters a realm that is physically located in the real world, makes use of props in the real world, is often about the real world, is said by the players to be real, and yet in some way is mentally removed from the real world" (Gray 2008, 6). The goals of play are typically unrelated to real or externally motivated consequences (Sapora and Mitchel 1961). In museums, exhibit phenomena often have real-world applications, but the consequences of an exhibit experience are usually far removed from real or serious life. As Frank Oppenheimer, physicist and founder of the Exploratorium once remarked, "No one ever flunked a museum" (as cited in Semper 1990). Unfortunately, adults sometimes eschew unserious play, looking for real-world relevance. According to Gray, adults experience play best when it includes motivations aligned with their mature responsibilities. While evaluation studies are not needed to assess an exhibit's imaginative, unserious nature, they may reveal successful designs for encouraging adults to play at exhibits.

Interview

Interviews may assess adults' needs for real-world applications or meaningful connections when engaging in play at exhibits. For example, interviews conducted at *Dividing Space* (where visitors rearrange pegs and roll pucks) suggested that adult visitors struggled to find a motivational purpose at the exhibit. When asked if there was anything confusing about the exhibit, several adults requested a connection to real-world motivations and responsibilities:

> *Adult male: I'm curious what's the meaningful use of it…?*
>
> *Adult female: [You] might want to explain the application of it, like how it is used in earth science. Like, what is it good for?*

The exhibit development team wished to maintain the playful core activity with pegs and pucks, but believed a stronger link to real-world application would assist in engaging adults. The solution was to create a secondary label linking the exhibit phenomenon to the real world; in this case, the label showed how cities use Voronoi diagrams to determine the placement of new fire stations as cities grow. In this example, the design-test cycle led to an exhibit label that successfully bridged adults' playful exhibit experience to real-world concerns and motivations.

Observation

Observations may effectively measure the range and type of visitor engagement present at exhibits. For example, at the *Gyroid*, visitors climb through a ten-foot high geometric structure with two separate networks of curving, intertwining tunnels (see Figure 7). As visitors crawl up and around, they experientially discover that the Gyroid is a two-room structure: the only way to get from one tunnel to the other is to leave the structure and reenter through a different opening. Engaging both children and adults on the play structure was an important project goal. Using real-time observations, we conducted "sweeps" of the area in and around the climber every 15 minutes, noting whether the adults and children in the space were reading labels, interacting with visitors on the climber, or climbing on or pointing to the structure. Observations indicated that only a few adults interacted with their children, and even fewer climbed inside the structure. To encourage greater adult engagement, the team developed a label with a direct invitation to adults: "Climb with your kid!" Again, we conducted observational sweeps. After the labels were added to the *Gyroid*, more adults were actively helping, guiding and climbing with their children. Conducting systematic observations helped create and assess a design that more effectively engaged adults in playful behavior.

Figure 7. *Gyroid* exhibit (photo courtesy of the Exploratorium)

Challenges of the Design-Test Cycle

Although the examples offered here may present the iterative design-test process as straightforward to implement, they gloss over several challenges. First, building and sustaining good working relationships between developers and evaluators may be difficult. At the Exploratorium, such relationships have been evolving for decades. As evaluators, we have employed two key strategies for sustaining strong relationships with exhibit developers: build trust and involve developers in the evaluation process (Gutwill 2010; Medlock et al. 2005). To build trust, we have focused on formative, rather than summative evaluation, so as to emphasize the immediate utility of evaluation studies (Patton 2008). Additionally, we approached only amenable developers at first, creating models of effective relationships before trying to persuade more hesitant developers to embrace evaluation. Finally, we have built trust by listening carefully to developers' concerns, even employing those concerns to hone research and evaluation questions. By pursuing questions developers raised, we hoped to deepen their involvement in the evaluation process. In addition, we brought developers into direct contact with visitors or visitor data by asking them to accompany us during interviews or observations, and by watching video together of visitors' interactions at exhibits. This shifted the evaluator out of the role of "translator" of visitors' experiences into "collaborator" with the developer to understand those experiences. Trust and involvement have strengthened relationships between developers and evaluators, but they require time and effort.

A second challenge in implementing a design-test process arises for institutions that do not create their own offerings (e.g., exhibits, programs, curricula), but adapt them from elsewhere. While these institutions may not be able to change the main design of an offering in response to formative evaluation, an iterative design-test process could be useful in adjusting some aspect during adaptation. For example, in institutions that do not build their own exhibits, formative evaluation could help improve interpretive labels, determine placement and orientation of exhibits, or design overall environments for exhibits. Changes to any of these may affect visitors' experiences at the exhibits. Indeed, much formative evaluation in museums has been conducted in these arenas (for early reviews, see Bitgood and Finlay 1986; Screven 1999).

The last challenge comes in the form of a lack of resources: Many institutions, even those that create their own offerings, believe they cannot afford to perform evaluation studies. We contend that "quick-and-dirty" formative evaluation can be conducted inexpensively, often producing useful information (Medlock et al. 2005). In museums, effects of exhibit (or other) changes on visitors' behaviors can often be detected by interviewing or observing a mere dozen visitors. Studies of small samples may be made more valid and incur little extra cost when evaluators (or other staff) act systematically, account for their biases, and record and analyze their data carefully and comprehensively. Many resources exist for staff with little training who wish to conduct formative evaluation studies (e.g., Allen 2000; Beaumont 2010; Diamond 1999; Friedman 2008; Giusti et al. 2009; Peterson 2004; Stevens et al. 1993; Taylor 1991).

Despite the challenges, we believe the rewards from implementing a process of iterative design and evaluation outweigh the costs. As evidenced above, the process often produces more playful and effective learning encounters in museums.

Conclusion

We have demonstrated how an iterative process of development and formative testing can improve museum exhibits, engendering more playful learning experiences for visitors. As researchers and evaluators, we focused on the evaluative part of the process, offering in detail two methods for gaining user feedback from visitors: interviews and observations. Using Peter Gray's five aspects of play as a framework, we provided examples in which these methods yielded information useful to our development team for creating more playful exhibit experiences. In *Geometry Playground,* the design-test cycles ultimately produced exhibits that:

- offered implicit constraints and rules for learners to internalize, thereby promoting targeted thinking skills such as spatial reasoning;
- allowed visitors to get started quickly and without frustration;
- supported process-oriented rather than goal-oriented activities;
- presented open-ended exhibit elements to communicate fewer stopping points and foster self-directed visitor exploration; and
- engaged both adults and children in playful learning experiences by inspiring imagination while connecting to real-world issues.

These features of improved exhibits map directly onto Gray's five aspects of play. While there are challenges to incorporating a design-test development process, we believe that it produces educational exhibits in which visitors feel and act more playfully. And playful learning experiences have serious benefits: attracting a broader range of learners, engaging them more deeply for prolonged periods of time, and empowering them to direct their own learning, finding enjoyment and satisfaction along the way.

Toni Dancu, *Ph.D., is a Researcher in the Department of Visitor Research and Evaluation at the Exploratorium. She applies her training in developmental psychology to studying the engagement of underserved audiences such as women and African Americans; the role of relevance in the museum experience; and the effects of exhibit characteristics on visitors' exhibit use, time spent, conversations, and more. Toni conducts formative evaluations to help create intuitive and engaging experiences that reflect exhibit developers' intentions by folding visitors' voices into the process.*

Josh Gutwill, *Ph.D., is the Director of Visitor Research and Evaluation at the Exploratorium. His work includes research on learning designs in informal educational environments as well as exhibit and program evaluation to improve visitors' experiences. He seeks to create opportunities for visitors to engage in self-directed inquiry in science museum settings. Another of his driving interests is to use research and evaluation methods to help educators (teachers, exhibit developers, curriculum designers) improve their practice.*

Nina Hido *is a Senior Project Evaluator at the Exploratorium. She has been involved in a range of front-end, formative, and research studies working closely*

with exhibit development teams and visitors on the museum floor. Nina is especially interested in how intuitive design can create accessible learning experiences.

References

Allen, S. (2000). "How is Writing a Good Set of Questions Like Designing a Good Exhibit?" *Visitor Studies Today!* 3(2): 17-18.

Allen, S., J.P. Gutwill, D. Perry, C. Garabay, K. Ellenbogen, J. Heimlich, C. Reich, and C. Klein (2007). "Research in Museums: Coping with Complexity." In J. Falk, L. Dierking, and S. Foutz, eds. *In Principle, In Practice: Museums as Learning Institutions*. New York: Rowman & Littlefield Publishers, Inc., 229-246.

Bakeman, R. (2000). "Behavioral Observation and Coding." In H.T. Reis and C.M. Judd, eds. *Handbook of Research Methods in Social and Personality Psychology.* Cambridge: Cambridge University Press, 138-159.

Bakeman, R., and J.M. Gottman (1997). *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge: Cambridge University Press.

Beaumont, L. "The Adult-Child Interaction Inventory: A Tool for Exhibit Development and Evaluation." Presented at Annual Meeting of the Association of Science-Technology Centers, Honolulu, HI.

Bitgood, S., and T. Finlay (1986). "Bibliography: Exhibit Signs, Labels and Graphics." *Visitor Behavior* 1(3): 6-7.

Cobb, P., J. Confrey, A. diSessa, R. Lehrer, and L. Schauble (2003). "Design Experiments in Educational Research." *Educational Researcher* 32(1): 9-13.

Csikszentmihalyi, M., and K. Hermanson (1995). "Intrinsic Motivation in Museums: Why Does One Want to Learn?" In J.H. Falk and L.D. Dierking, eds. *Public Institutions for Personal Learning: Establishing a Research Agenda*. Washington, DC: American Association of Museums, 67-77.

Derry, S.J., R.D. Pea, B. Barron, R.A. Engle, F. Erickson, R. Goldman, R. Hall, T. Koschmann, J.L. Lemke, M.G. Sherin, and B.L. Sherin (2010). "Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology, and Ethics." *Journal of the Learning Sciences* 19: 3-53.

Diamond, J. (1999). *Practical Evaluation Guide: Tools for Museums and Other Informal Educational Settings*. Walnut Creek: AltaMira Press.

Falk, J.H. (1983). "Time and Behavior as Predictors of Learning." *Science Education* 67(2): 267-276.

Falk, J.H., and Dierking, L.D. (2000). *Learning from Museums: Visitor Experiences and the Making of Meaning*. Walnut Creek, CA: Altamira Press.

Friedman, A.J. (2008). *Framework for Evaluating Impacts of Informal Science Education Projects*. Washington, D.C.: National Science Foundation.

Giusti, E., S. Yalowitz J. Koke, and C. Garibay (2009). "Visitor Studies 101 Handout." Presented at American Association of Museums Conference, Philadelphia, PA.

Good, C.V., ed. (1973). *Dictionary of Education, 3rd ed*. New York: McGraw-Hill.

Gray, P. (2008). "The Value of Play I: The Definition of Play Provides Clues to Its Purposes." Psychology Today: Freedom to Learn (March 1). Available from: http://www.psychologytoday.com/blog/freedom-learn/200811/the-value-play-i-the-definition-play-provides-clues-its-purposes

Gutwill, J.P. (2003). "Gaining Visitor Consent for Research II: Improving the Posted-Sign Method." *Curator* 46(2): 228-235.

----- (2005). "Observing APE." In Humphrey, T. and J. Gutwill, eds. *Fostering Active Prolonged Engagement: The Art of Creating APE Exhibits*. Walnut Creek: Left Coast Press.

Gutwill, J.P. (2010). "Promoting Participatory Evaluation at the Exploratorium." Presented at the Annual Meeting of the Visitor Studies Association, Phoenix, AZ.

Humphrey, T., and J.P. Gutwill (2005). *Fostering Active Prolonged Engagement: The Art of Creating APE Exhibits*. Walnut Creek: Left Coast Press.

Hutt, C. (1981). "Toward a Taxonomy and Conceptual Model of Play." In Day, H.I. ed. *Advances in Intrinsic Motivation and Aesthetics*. New York: Plenum Press, xii, 503.

Leinhardt, G., and K. Crowley (1998). *Museum Learning as Conversational Elaboration: A Proposal to Capture, Code, and Analyze Talk in Museums*. Pittsburgh, PA: Learning Research & Development Center at University of Pittsburgh.

Linn, M.C., and A.C. Petersen (1985). "Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis." *Child Development* 56(6): 1479-1498.

Medlock, M.C., D. Wixon, M. McGee, and D. Welsh (2005). "The Rapid Iterative Test and Evaluation Method: Better Products in Less Time." In Randolph, G., D. Bias, and J. Mayhew, eds. *Cost-Justifying Usability: An Update for an Internet Age*. San Francisco: Morgan Kaufman.

Meisner, R. (2007). *Encounters with Exhibits: A Study of Children's Activity at Interactive Exhibits in Three Museums*. Dissertation, Kings College, London.

Mornighan-Nourot, P., B. Scales, J. Van Hoorn, and M. Almy (1987). *Looking at Children's Play: A Bridge between Theory and Practice*. New York: Teachers College Press.

National Research Council (2009). *Learning Science in Informal Environments: People, Places, and Pursuits*. Washington, D.C.: National Academies Press.
----- (1996). *National Science Education Standards*. Washington, D.C.: National Academy Press.

Norman, D. (1988). *The Design of Everyday Things*. New York: Doubleday Currency.

Oppenheimer, F. (1972). "The Exploratorium: A Playful Museum Combines Perception and Art in Science Education." *American Journal of Physics* 40(July): 978-984.

Packer, J. (2006). "Learning for Fun: The Unique Contribution of Educational Leisure Experiences." *Curator* 49(3): 329-334.

Patton, M.Q. (2008). *Utilization-Focused Evaluation*. Los Angeles: Sage.

Perry, D. (1993). "Designing Exhibits That Motivate." In McNamara, P., M. Borun, S. Grinell, and B. Serrell, eds. *What Research Says about Learning in Science Museums*. Washington, D.C.: Association of Science-Technology Centers.

Peterson, A. (2004). *W.K. Kellogg Foundation Evaluation Handbook*. W.K. Kellogg Foundation.

Piaget, J. (1953). *The Origin of Intelligence in the Child*. London: Routledge & Kegan Paul Ltd.

Rathunde, K., and M. Csikszentmihalyi (2006). "The Developing Person: An Experiential Perspective." In Damon, W., R.M. Lerner, and N. Eisenberg, eds. *The Handbook of Child Psychology*. Hoboken, NJ: John Wiley & Sons, Inc.

Rosenthal, R., and R.L. Rosnow (2008). *Essentials of Behavioral Research: Methods and Data Analysis, 3rd ed.* Boston: McGraw Hill.

Sapora, A.V., and E.D. Mitchel (1961). *The Theory of Play and Recreation*. New York: The Rondald Press Company.

Sauber, C.M., ed. (1994). *Experiment Bench: A Workbook for Building Experimental Physics Exhibits*. St. Paul: Science Museum of Minnesota.

Screven, C.G., ed. (1999). *Visitor Studies Bibliography and Abstracts, Fourth ed.* Chicago: Screven & Associates.

Semper, R.J. (1990). "Science Museums as Environments for Learning." *Physics Today* 43(11): 50-56.

Serrell, B. (1998). *Paying Attention: Visitors and Museum Exhibitions*. Professional Practice Series, American Association of Museums.

Smith, R., C. McPhail, and R. Pickens (1975). "Reactivity to Systematic Observation with Film: A Field Experiment." *Sociometry* 38(4): 536-550.

Stevens, F., F. Lawrenz, and L. Sharp (1993). *User-Friendly Handbook for Project Evaluation: Science, Math, Engineering and Technology Education*. Washington, D.C.: National Science Foundation.

Tarte, L. (1990). "Spatial Orientation Skill and Mathematical Problem Solving." *Journal for Research in Mathematics Education* 21(3): 216-229.

Taylor, S., ed. (1991). *Try It! Improving Exhibits through Formative Evaluation*. New York: New York Hall of Science, ASTC.

Vissner, P.S., J.A. Krosnick, and P.J. Lavrakas (2000). "Survey Research." In Reis, H.T. and C.M. Judd, eds. *Handbook of Research Methods in Social and Personality Psychology*. Cambridge: Cambridge University Press, 223-252.

Vygotsky, L.S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.