Facilitating Family Group Inquiry at Science Museum Exhibits

JOSHUA P. GUTWILL
Exploratorium, 3601 Lyon Street, San Francisco, CA 94123, USA

SUE ALLEN
Exploratorium, 3601 Lyon Street, San Francisco, CA 94123, USA, and
National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230, USA

Received 13 February 2009; revised 5 November 2009; accepted 11 November 2009

DOI 10.1002/sce.20387
Published online 28 December 2009 in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: We describe a study of programs to deepen families' scientific inquiry practices in a science museum setting. The programs incorporated research-based learning principles from formal and informal educational environments. In a randomized experimental design, two versions of the programs, called *inquiry games*, were compared to two control conditions. Inquiry behaviors were videotaped and compared at pretest and posttest exhibits. Family members were also interviewed about their perceptions and use of the inquiry games. Results indicated that visitors who learned the inquiry games improved their inquiry more than those who did not. Effect sizes ranged from 0.3σ to 0.7σ, depending on the assessment measure. Visitors who learned the collaborative inquiry game showed even more improvement than those who learned the individualized game, spending more time investigating the posttest exhibit, making more frequent and more abstract interpretations of their experiments, building more collaborative explanations, and engaging in more coherent inquiry investigations than controls. Qualitative analysis suggested that the collaborative inquiry game was superior because it required all family members to participate, work

Correspondence to: Joshua P. Gutwill; e-mail: jgutwill@exploratorium.edu
Contract grant sponsor: National Science Foundation.
Contract grant number: 0411826.
This work supported by the Foundation was done while the second author was working at the Foundation.
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

together, and explicitly articulate their interpretations. Visitors in all conditions enjoyed their experience, varied in what they liked and disliked, and reported applying what they had learned at new exhibits. *Sci Ed* **94:**710–742, 2010

INTRODUCTION

Scientific inquiry skills, such as questioning, observing, predicting, experimenting, explaining, and communicating ideas to others, are widely seen as foundational to science learning in and out of schools (e.g., Adey & Shayer, 1993; Ansbacher, 1999; Ash, 1999; Chinn & Malhotra, 2002; Hestenes, 1987; Kuhn & Udell, 2003; National Research Council [NRC], 2007, 2009; Tien, Rickey, & Stacey, 1999; B. White, 1993; B. White & Frederiksen, 2000; R. T. White, 1988). The National Science Education Standards (NRC, 1996, 2000) argue that engaging students in inquiry helps them to understand science concepts and the nature of science, as well as to develop skills to become independent inquirers about the natural world.

However, most inquiry-based learning materials in science have been developed for school settings. By contrast, we argue that designed informal settings such as science museums constitute ideal environments for teaching and learning inquiry skills, offering a number of features often unavailable in schools. First, science museum settings offer direct access to a vast array of phenomena in the form of exhibits; many of these are designed to support investigation and demystify science (Oppenheimer, 1968, 1972; Semper, 1990). Second, casual museum visitors engage in authentic, self-directed inquiry, learning not for external rewards but because the activity itself is sufficiently engaging to sustain their interest (Csikszentmihalyi & Hermanson, 1995; Falk & Dierking, 2000). Third, learning in museums is a fundamentally social experience, often taking place in multigenerational groups that draw on each member's interests and expertise as the group interacts with exhibits (Falk, Moussouri, & Coulson, 1998; Hilke, 1985). Finally, museums are widely available resources for lifelong learning (e.g., Ansbacher, 1999; Bell, Lewenstein, Shouse, & Feder, 2009; Hein, 1998), and their emphasis on hands-on interactivity and playful exploration offers an opportunity to involve broad audiences who would not consider themselves as having knowledge or even interest in science.

At the same time, research has shown that the majority of casual visitors do not make full use of these opportunities to engage in deep, extended inquiry. For example, Randol (2005) found that the most common inquiry behaviors at a sample of interactive exhibits were "do and see" (i.e., manipulate and observe the outcome), whereas more advanced inquiry strategies such as describing results or drawing conclusions were relatively rare. Allen (2002) found that prediction and metacognition were likewise unusual in visitors' conversations at a science museum exhibition on frogs, present in less than 10% of the "learning talk." Crowley and colleagues (Crowley & Galco, 2001; Crowley & Jacobs, 2002) showed that visitors' explanations at exhibits in a children's museum tend to be short and isolated "explanatoids" rather than extended or chained inferences. Numerous studies have shown that visitors generally use interactive exhibits for short times, of the order of 1 minute or less (e.g., Diamond, 1986; Gutwill, 2005), and museum staff frequently look for ways to "slow people down."

Many factors probably contribute to this underutilization of museums as a powerful resource for deep, authentic inquiry. The museum environment itself, with its typically open floor plan and myriad visual distractions, can undermine the focus required for sustained inquiry (Adamson, 2008; Allen, 2004c). Another environmental problem is overcrowding; visitors waiting for an exhibit may pressure current users to end their inquiry and relinquish the exhibit. Exhibit designs themselves may limit opportunities for inquiry; some do not offer enough variables for visitor manipulation, forcing visitors into a predictable sequence of "do, notice, read" behavior (Gutwill, 2008; Tisdal & Perry, 2003).

To address such issues, museums have explored a range of design strategies in recent years, including designing exhibits that present multiple options for visitor manipulation (Borun et al., 1998; Humphrey & Gutwill, 2005), and "multiple-station" exhibits that can be used by different visitors simultaneously, reducing the pressure to move on because of crowding (Gutwill, 2005). Yet, even at these "best cases" of exhibit design, few visitors engage in deep inquiry at exhibits for longer than a few minutes. One possible explanation is that visitors lack the skills in inquiry that would allow them to deeply investigate the phenomena at such exhibits. For example, Allen's study of the popular Colored Shadows exhibit (1997) showed that fewer than 10% of visitors were able to create simple experiments that would discriminate among models or to revise a model in light of contradictory evidence. This suggests that some inquiry skills may be exceptionally difficult for visitors to employ, even when directly asked to do so.

It seems likely that human mediation may be key to helping museum learners take full advantage of their environment (NRC, 2009). Science museums have long relied on informal docents to help guide visitors' experimentation, and some are extending these efforts into more elaborate public programs. For example, public programs by the Play, Invent, Explore (PIE) Institute in science museums combine inquiry processes with design challenges to foster sustained visitor exploration (St. John, Carroll, Helms, & Smith, 2008). Much of the work in this area has actually been conducted in art museums (e.g., Davis, 1993; Housen, 2001). While such efforts have shown great promise, little is known about the specific characteristics of programs that effectively support both personal and group inquiry learning at science museum exhibits. Even less is known about how inquiry-based science museum programs may help visitors build skills they can apply at new exhibits, outside the program itself.

This challenge—the creation and study of public programs to teach museum visitors an effective inquiry process—was undertaken in the Group Inquiry by Visitors at Exhibits (GIVE) project at the Exploratorium, a science museum in San Francisco, California. Because of the uniqueness and complexity of museums as free-choice learning environments, one could not expect to simply apply classroom programs to museums. We assumed that a successful intervention would be grounded in cognitive science and sociocultural theory and respond to visitors' personal experiences and search for meaning. Moreover, the inquiry approach taught in science museum programs would have to be brief, memorable, useful, and appealing enough that visitors would grasp and apply it at novel exhibits.

RESEARCH QUESTIONS

Given the current status of museums as a potentially powerful yet underutilized resource for genuine, sustained inquiry by the public, the central questions driving the GIVE research project were as follows:

- Can intergenerational groups of museum visitors, such as families, be coached by museum staff to learn a set of inquiry skills that they can use on their own at novel exhibits or even experiences beyond their visit?
- How does such an intervention affect visitors' inquiry behaviors at a novel exhibit; can it support them to explore the exhibit more deeply?
- What properties of the staff-mediated intervention seem central to its design, and what is the evidence for them?

To answer these questions, the GIVE team created and iteratively studied two *inquiry games*, public programs in which groups of visitors learned simple inquiry skills they could use as they interacted with exhibits. The process for developing the inquiry games incorporated design-based research in its early stages, culminating in a rigorous randomized experimental study. Details of our formative process for producing the games are published elsewhere (Allen & Gutwill, 2009).

DESIGN OF THE INQUIRY GAMES

As mentioned previously, inquiry-based curricula in schools often focus on a range of skills, such as questioning, predicting, designing an experiment, observing, analyzing, communicating, and so on. In the museum setting, constraints of timing and cognitive load forced us to gradually reduce the number of skills to those that families could learn and recall at novel exhibits.

Inquiry Skills

In their final form, the inquiry games focused on the following two skills:

1. *Proposing Actions* (PA). This skill involves asking a question or making a plan at the beginning of an investigation.
2. *Interpreting Results* (IR). This includes making observations, interpretations, or explanations during or after an investigation.

These two skills complement visitors' spontaneous physical experimentation at exhibits and were intellectually accessible to diverse groups. They were simple enough to be understood quickly and remembered easily. Moreover, research has shown that these skills are not routinely practiced by visitors (Randol, 2005). Even if visitors privately ask questions or draw conclusions as they use exhibits, articulating such thoughts publicly is a notable inquiry skill in itself and moves the family's investigation forward by putting ideas into public spaces for recognition and improvement by peers (Quintana et al., 2004; Scardamalia, 2002). We further hoped that these skills might act as gateways to other skills, such as explanation building or hypothesis testing.

Our inquiry process, as shown in Figure 1, allowed visitors to cycle through its steps several times at an exhibit. This mirrors the cyclical format found in many inquiry-based

Figure 1. Simplified inquiry cycle with the two targeted skills shown in black.

curricula in schools, where experiments and conclusions lead to new questions (e.g., Champagne, Kouba, & Hurley, 2000; Chinn & Malhotra, 2002; Songer, 2004; B. White & Frederiksen, 1998). The dark ovals in Figure 1 indicate the skills we explicitly taught visitors, whereas the light oval represents physical experimentation, a skill visitors spontaneously practice when using an interactive science museum exhibit.

Pedagogical Structures

In developing the inquiry games, we had to create appropriate activity structures within which the two inquiry skills could be used at an exhibit. Based on extensive pilot testing, the team settled on two different inquiry games, one structured and collaborative and the other more supportive of individual spontaneity.

Juicy Question Game. In the first game, called Juicy Question, the family works together to identify and jointly investigate a single question that is "juicy," explicitly defined as a question that can be answered at the exhibit and to which nobody knows the answer. To begin, visitors explore the exhibit to become familiar with it. Then, family members take turns sharing a question they each have about the exhibit. Next, they choose one of the questions to investigate as a group and conduct their experiment. Finally, they stop to reflect on what they have discovered, sharing their ideas until they feel finished. Often, their ideas lead to new questions and experiments. Each family member receives a card as a reminder of the two skills of asking juicy questions (PA) and sharing discoveries (IR).

Because the Juicy Question game is highly collaborative, it requires a facilitator. At first, the museum educator takes this role but then asks one of the adults in the family to gradually take it over. The facilitating adult is also given a special card with steps for running the game, to serve as a reminder and support. Each family plays the game at two different exhibits before being asked to do it without the museum educator at the posttest exhibit (see Figure 2 for the experimental design).

Hands-Off Game. The team was concerned that the highly structured nature of Juicy Question might make it seem less like a game and more like formal schooling. For this reason, we developed an alternative activity structure, which was more spontaneous, in keeping with the informal and choice-based environment of the museum floor. The "Hands-Off" game allows anyone, at any time, to call out "hands off!" at which time the other family members must stop using the exhibit and listen to the caller. The caller can share either a proposal for something they wish to investigate (PA), or a discovery about the exhibit (IR), before calling "hands on" again. Thus, the game teaches visitors to use the same two inquiry skills as Juicy Question, but in a more individual and spontaneous way.[1] No facilitator is needed, and the game is particularly easy to remember. For consistency with Juicy Question, one adult was given a special card reminding them to support the rules of the game. In our pilot study, Hands-Off proved to be especially popular with younger children, since it gave them a way to be heard over their potentially dominating siblings.

[1] The Hands-Off game was partly inspired by the Visual Thinking Strategies (VTS) approach taken in art museums (Housen, 2001; Housen & Yenawine, 2001). Like VTS, Hands-Off encourages group members to focus on interesting phenomena, engage in a spontaneous participation structure, and repeat a set of probes to make interpretations of the phenomena, all of which result in the emergence of a shared understanding of the exhibit.

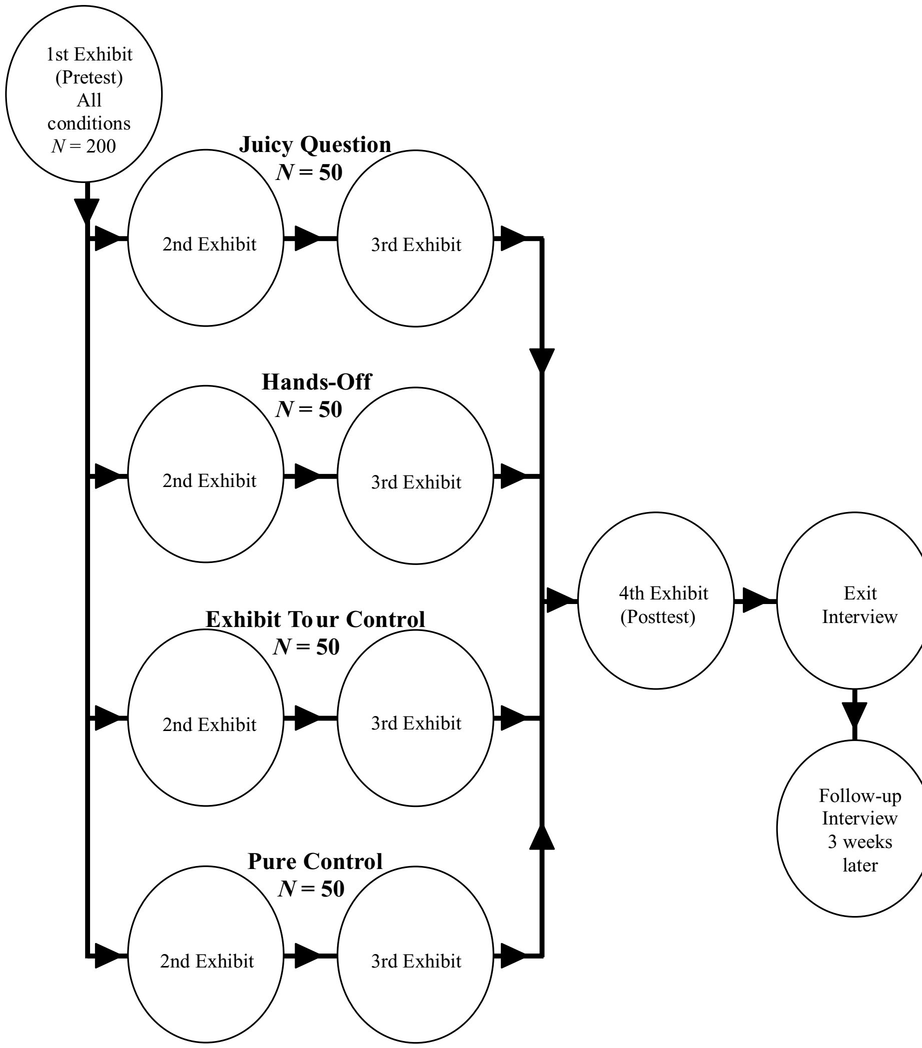

Figure 2. Experimental design and flow of participants.

Principles Underlying the Inquiry Games

Regardless of their particular pedagogical structure, both inquiry games incorporated several key principles from the school-based science education literature:

Build on Learners' Prior Knowledge. In their review of learning research, Bransford, Brown, and Cocking (2003) stress the importance of viewing humans as "goal-directed agents who actively seek information" (p. 10), building new understandings on prior knowledge (also Piaget, 1978; von Glasersfeld, 1991; Vygotsky, 1962). In both inquiry games, we incorporated this principle by letting families explore an exhibit and gradually surface their own questions about it, rather than prescribing specific science content that should be learned, irrespective of learners' ages and experiences.

Teach via Modeling, Scaffolding, and Fading. We embraced the well-studied pedagogical approach of "cognitive scaffolding" (Wood, 2001), in which an educator uses questions, prompts, and other structured interactions as cognitive supports for learners during an extended investigation (Samarapungavan, Mantzicopoulos, & Patrick, 2008). The educator then gradually "fades," allowing the learners to continue autonomously (Collins, Brown, & Newman, 1989; Vygotsky, 1978). In the current study, families played an inquiry game at three novel exhibits. The educator gradually faded from the role of facilitator, first by helping one of the adult visitors facilitate, then by disappearing and giving the family full autonomy.

Identify Skills Explicitly. Numerous studies in the learning sciences have shown that people learn new skills better if it they are explicitly articulated, demonstrated, and practiced (Labudde, Reif, & Quinn, 1988; Palincsar & Brown, 1984; B. White & Frederiksen, 1998). In our case, the educator explicitly stated the purpose and key steps of the inquiry game and named instances of the specific skills as visitors practiced them in the context of the conversation at the exhibit (e.g., "Yes, that sounds like a discovery," or "Does anyone else have a question they want to ask?").

Support Metacognition. Another key principle from formal schooling is supporting metacognition, learners' abilities to monitor and reflect on their own learning (Bransford et al., 2003; Brown, 1975; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Flavell, 1973). In this study, we targeted two skills that had a strong metacognitive component: asking a question and reflecting on what was learned during an investigation. Both skills require the learner to monitor their own knowledge state and its changes.

Support Collaboration. Research has shown that learning in groups often increases motivation as well as achievement (Tobin, Tippins, & Gallard, 1994). Our entire study took as its participants only groups (typically multigenerational families) and relied on their existing social dynamics and established strategies for learning. In fact, this study directly investigates the effect of group collaboration on inquiry by comparing a highly collaborative inquiry game (Juicy Question) with a more individualized one (Hands-Off).

In addition to these adaptations of school-based principles for inquiry, we added several other design principles based on the research on learning in informal settings.

Make the Activity Intrinsically Motivating. Learning in museums is self-directed and voluntary (Allen, 2004a; Falk & Dierking, 2000; Hein, 1998), so we expected visitors to use the inquiry skills only if they were in a format that was intrinsically motivating to them and did not require extended or effortful learning. We kept the entire session brief (typically, 10–15 minutes per exhibit) and adjusted the skills and pedagogical structure to optimize visitors' self-reported enjoyment and satisfaction, based on Perry's (1993) components of a motivating exhibit experience: curiosity, confidence, challenge, control, play, and communication.

Minimize Cognitive Load. To ensure that visitors could easily remember the skills we had taught them, we reduced the number of skills from the six we originally envisioned down to two. To further support visitors' recall capacities, we handed them glossy, double-sided cards that summarized the inquiry skills with colorful icons and key phrases (adaptation of Davis, 1993).

Support Visitors' Learning Agendas. Visitors have their own agendas in museums (Doering, 1999; Ellenbogen, 2002; Falk, 2009; Falk et al., 1998; Hilke, 1985), so the elements of the inquiry games were adjusted to accommodate visitors' spontaneous behavior, expectations, and identity-building activities. For example, some parents preferred to hang back and let their children guide the activity, whereas others had a teaching agenda. We allowed this kind of flexibility where possible, to maximize the chances that the family would spontaneously use the inquiry skills beyond the training period.

Support Individual Spontaneity. One of the games (Hands-Off) was designed to emphasize personal choice and spontaneous participation, in keeping with the self-directed and voluntary nature of the learning setting.

STUDY DESIGN

Following the design-based development of the inquiry games, we conducted a randomized controlled study to test their impact on families' inquiry behaviors at exhibits. We randomly assigned participating families to one of four conditions and compared their behaviors while using the same set of exhibits. The four conditions differed in terms of the kind of guidance or mediation participants received while at the exhibits:

1. *Juicy Question*: Families learn and play the "Juicy Question" game with a staff educator.
2. *Hands-Off*: Families learn and play the "Hands-Off" game with a staff educator.
3. *Exhibit Tour Control*: Families receive an interactive description from a staff educator about the development history and science content of the exhibits. Although families do engage with the exhibits in this condition, the educator is careful not to model or teach inquiry.[2] The goal was for families to have an enjoyable, interesting interaction with an educator without learning how to do inquiry.
4. *Pure Control*: Families use the exhibits without any game or educational mediation. The purpose of this condition was to control for practice effects, in case families get better at doing inquiry simply by using exhibits.

We studied 50 families in each of the four conditions, 200 families altogether.[3] We recruited families from the museum floor[4] and conducted the study in a research laboratory at the back of the museum, so we could videotape families using exhibits for detailed analysis. By comparing families who had learned the inquiry games to families who had not, we could determine whether the games helped families conduct in-depth inquiries at exhibits.

[2] The Exhibit Tour was designed only as a control for the effect on family inquiry of working with an educator, rather than as an example of best practice in content-based tours (theoretically and logistically beyond the scope of this study).

[3] While we refer to them as "families," there was no strict requirement for their relationships to each other; we simply recruited visiting groups who were not members, with at least one adult, at least one child aged 8–13, and no children younger than 8 years old.

[4] For participating, families were offered a thank-you gift worth approximately $20 from the museum store.

Hypotheses and Planned Comparisons

Before collecting our data, we articulated our hypotheses for the study and determined the planned comparisons we would make to test those hypotheses. We had three overarching hypotheses, with concomitant planned comparisons:

1. *Inquiry games will improve inquiry behaviors.* We predicted that families who learned the inquiry games would use the inquiry skills we taught them (PA and IR) more frequently and for longer durations than control group families. We thought it possible but less likely that they would also outperform the control group families in their frequencies of other inquiry behaviors that were not directly targeted (see the section Inquiry Codes). Overall, our planned comparison predicted an *effect of inquiry* that the Juicy Question and Hands-Off conditions would outperform the Exhibit Tour Control condition.
2. *Mediation will enhance the experience*. On the basis of studies showing that interactions with museum staff are among the most memorable aspects of a museum experience (e.g., Allen, 2004b; Piscitelli & Weier, 2002), we predicted that families who spent time with our engaging and knowledgeable museum educator would be inspired and encouraged to spend more time on their own with a new exhibit, perhaps resulting in deeper inquiry. Thus, our planned comparison predicted an *effect of mediation*, that the Exhibit Tour Control condition would outperform the Pure Control condition.
3. *Pedagogy will affect inquiry.* We had competing hypotheses about which pedagogical approach—group collaboration or individual control—would be more successful at helping visitors engage in inquiry. We predicted that the collaborative nature of Juicy Question would make for deeper inquiry, whereas the individualized character of Hands-Off would make it easier for families to use and adopt. Consequently, our planned comparisons employed two-tailed measures to test for an *effect of pedagogy*.

Data Collection

The sequence of participation for each family was as follows: They entered the laboratory, used a first exhibit as they normally would (pretest), learned to play one of the inquiry games at two more exhibits (unless they were in one of the two control conditions), and used a final exhibit on their own (posttest). At the final exhibit, we asked them to play the inquiry game if they had learned one, or simply use the exhibit normally if they were in one of the control groups. Finally, one adult and one child were randomly selected to participate in interviews after the experience (see Figure 2).

Exhibits in the Study. The inquiry games were designed to function at a broad range of exhibits, helping visitors investigate any phenomena presented. The particular exhibits used in the study did not focus on a shared science topic, but were originally developed to allow for deeper experimentation, and had several inquiry friendly qualities:

- *open ended*, having no obvious stopping point or “answer” to convey,
- *multioption,* offering numerous, related options for visitors to explore, and
- *multiuser*, permitting several visitors to engage simultaneously by providing different points of access.

Figure 3. A family uses the four exhibits in the study. The exhibits, clockwise from top left, are Shaking Shapes (pretest), Floating Objects, Unstable Table, and Making Waves (posttest).

Previous studies have found that these exhibit characteristics often lead to more visitor-driven investigation (Borun et al., 1998; Humphrey & Gutwill, 2005). A brief description of each of the exhibits employed in the study follows. The exhibits are shown in Figure 3.

Pretest Exhibit. *Shaking Shapes* is a low, vibrating table supporting a large set of variously shaped, loose geometric tiles that move in dance-like patterns—rotating, migrating, or falling. Shaking Shapes offers visitors the opportunity to assemble and test the stability of various geometric structures.

Second Exhibit. *Floating Objects* allows visitors to levitate a variety of objects on two fully adjustable blowers, experimenting with the height, rotation, and stability of the objects. By angling the blowers, visitors can investigate the effects of a nonvertical air stream.

Third Exhibit. *Unstable Table* challenges visitors to build block structures on a gimbaled table. By maintaining balance and building bigger structures, visitors may explore the effects of counterweighting, gravity, and torque.

Posttest Exhibit. *Making Waves* is a wave machine consisting of 20 magnetic pendulums hanging from a single spine. Visitors can experiment with the feedback from coupled magnets to explore interacting wave patterns.

Exhibit Sequence. Families in all conditions used the four exhibits in the fixed sequence described above, with Shaking Shapes serving as the pretest and Making Waves the posttest. We considered a design that counterbalanced the pretest and posttest exhibits to assess interaction effects between exhibit sequence and treatment condition. However, we rejected this design for two reasons. First, the same degree of analytic power in a crossover design would have required an increase in sample size that our resources did not permit. Second, the data collection process was already complicated, blocking for two variables (educator and condition) as well as utilizing blind recruiters, so an additional blocked variable might have led to considerable human error in administering the study. The final design maximizes power for identifying the differential treatment effects of our program, with the limitation that it may not generalize beyond this particular exhibit sequence.

Assessing Inquiry

Using the video data, we assessed families' inquiry behaviors at the first and last exhibit to determine whether playing the games had enhanced their inquiry skills. (We did not rigorously code families' behaviors at the intermediate exhibits because of resource constraints.) In practical terms, assessing families' inquiry behaviors involved creating several dependent variable codes and checking their reliability across multiple coders.

Inquiry Codes. To assess inquiry, we created the following codes.

Engagement. We measured the length of time a family chose to spend at the exhibit (i.e., holding time) as an indicator of their engagement. This was measured as the time from the first video frame after the educator finished her introduction to the last frame in which someone from the family touched or looked at the exhibit.

Proposing Actions (Spontaneous Use of Skill 1). We captured both the frequency and duration of PA utterances by families, which was one of the skills targeted in the intervention. Both the Juicy Question and Hands-Off games taught visitors to propose actions aloud when using exhibits. Proposed actions could take the form of a question or a statement. The PAs were coded at two levels—low and high. A low-level PA (PA1) would include only the action a visitor wanted to take ("What if we push this one?") or only the effect they hoped to induce ("Can we get all the magnets to swing together?"), but not both. A high-level PA (PA2) would include both the action and the desired result ("Let's see if we can get the middle magnets to swing hard by only pushing the end magnets").

Interpreting Results (Spontaneous Use of Skill 2). We counted both the frequency and duration of IR instances. Both games explicitly encouraged families to interpret the results of their experiments (e.g., "That was really fast"). As with PAs, we coded IRs at two levels. A low-level IR (IR1) would emphasize direct observation, with little generalization beyond the moment ("The red ones aren't moving"). A high-level IR (IR2) would include generalization ("If you move one, it stays in a pattern better than if you move more than one"), analogy, or explanation ("Those shapes fell first because their center of gravity was higher").

Collaborative Explanations. We counted the number of times multiple family members interpreted results in rapid succession and called these "consecutive IRs." Typically such conversational turn taking indicated that family members were building explanations jointly with others in a collaborative process. An example would be "That was really fast" followed immediately by "Yeah, because the faster you move the pendulum, the faster the wave travels." Consecutive IRs often occurred in strings of various lengths. For this code, we counted the percentage of a family's IRs that were consecutive, as well as the longest string of consecutive IRs spoken by a family.[5]

[5] For two IR utterances to be counted as "consecutive," the gap between the end of the first utterance and the start of the second had to be less than 2 seconds. This ensured that two different people uttered the two IRs.

Coherent Investigations. To assess the "depth" of a family's inquiry, we counted the number of times a family connected their investigations so that one experiment followed another in a coherent way (e.g., "Let's try these two and see how fast the wave goes," followed by "Now let's try this one and see if that makes it faster"). This meant looking at each PA to see whether it was linked to the one before it by a common theme (e.g., wave speed). We counted the number of PAs that were linked in this way and called the longest string of them the "linked PA" score. This was not a skill targeted in the games but was taken as an additional and independent indicator of the overall quality of the family's inquiry.

Inquiry Coding Scheme Reliability. Multiple research assistants, blind either to condition or to the purpose and hypotheses of the study, coded the videos of families at the pretest and posttest exhibits. To measure the reliability of our video-coding schemes, we randomly selected 16% of the videos across condition for coding by two research assistants, then checked their codes for interrater agreement.

The measures for reliability were chosen to be appropriate to each dependent variable code. For the PA and IR codes, we calculated the Cohen's kappa statistic for interrater agreement, a measure that accounts for the probability of agreements happening by chance (Bakeman & Gottman, 1997). The kappa statistic for these codes was found to be 0.76. Although there is no universal benchmark in the literature for acceptable levels of agreement, a kappa greater than 0.75 is considered to be "good" to "excellent" agreement beyond chance (Fleiss, Levin, & Paik, 2004; Landis & Koch, 1977). For the Linked PA codes (coherent investigations), we calculated Cronbach's generalizability coefficient, which assesses the degree to which different coders distinguish among families in similar ways (Bakeman & Gottman, 1997). The coefficient for Linked PA codes was found to be 0.88, which exceeds the convention for acceptable reliability (Hollander & Harvey, 2002). The scores for Consecutive IRs (collaborative explanations) were calculated automatically using the IR data, so did not necessitate an additional check for interrater reliability. Likewise, the interrater reliability was not necessary for time at exhibit (engagement), because frame-by-frame measurements were easily accurate to within 1 second.

Participants' Self-Reports

In addition to capturing and analyzing videotape data, we conducted interviews and surveys to assess demographics, visitors' satisfaction with the experience, and visitors' behaviors after the experience. Two types of interviews were administered. Immediately after the experience of using the exhibits, we randomly selected and interviewed one adult and one child from each family. These interviews focused on what they liked and disliked about the experience and, if applicable, about the game they learned. Participants were also asked about whether they thought they would use the game (or the Exhibit Tour information) in the future and whether we should offer it to other visitors. Three weeks later, we interviewed the same adult and child again by phone, asking what they could remember of the experience, and whether they used the game (or tour information) on the museum floor or outside the museum. All interview data were coded redundantly by two research assistants who negotiated any disagreements, unless otherwise noted below.

A demographic survey, administered in the laboratory, asked all family members about their age, gender, highest level of schooling, and highest level of science attained in school. The survey data were used in the analysis both to ensure that there were no demographic differences across treatment and control conditions and to determine whether demographic variables interacted with the treatments.

Assessing Conceptual Understanding

Although the program was developed to deepen scientific inquiry rather than teach any specific content, we were interested to see whether families' interpretations of their own investigations were generally correct or incorrect, with respect to canonical science. In keeping with the informal nature of learning in museums, we did not wish to "test" families in the interview, so we indirectly measured correctness of understanding by revisiting the video data and recoding all high-level interpretations (IR2s) in terms of their fit with scientific canon. Two coders separately categorized each IR2 as correct, incorrect, or uncodable, then met to work out all disagreements. An "uncodable" response would be one in which the coders (a) could not understand the words spoken, (b) could find no science content relevant to the exhibit, or (c) could not categorize the utterance as "correct" or "incorrect" within five viewings of it.

Statistical Analyses

To implement our planned comparisons in statistical tests, we used a pre/post difference score for each dependent variable coded from the videotapes. Interview and survey data yielded only posttest scores. The difference scores or posttest scores were then compared across conditions.

RESULTS

Overall, our planned comparisons revealed that the inquiry games increased the quantity and quality of families' scientific inquiry from pretest to posttest in several ways. Further analysis indicated that the Juicy Question game was mainly responsible for the improvement in visitors' inquiry behaviors. In this section, we first report on the demographics of the families in the different conditions. Next, we describe the effects of the inquiry games on visitors' inquiry behaviors. We then report visitors' reflections on their experience. Finally, we share the results of our modest findings on the correctness of families' interpretations of exhibit phenomena.

Demographic Attributes of the Families

Chi-square tests of the demographic survey data indicate that there were no differences by condition in any of the following:

- age of oldest and youngest child/age of oldest adult,
- highest grade attained by a child/highest level of schooling attained by an adult,
- highest level of science schooling attained by a child/by an adult,
- number of children with "special knowledge, interest or training in science"/number of adults with "special knowledge, interest or training in science," and
- number of girls, women, and men in the family.

Only one demographic variable showed a difference: There were significantly fewer boys in the Juicy Question families than in the Hands-Off families (planned comparison $\chi 2 = 9.2$, $p < .01$). By chance, more families with few boys were assigned to the Juicy Question condition. However, further analysis suggested that this did not significantly change the pattern of results.[6]

Although we found few differences in demographics across conditions, we hoped to use the demographic variables of age and education level as covariates in our planned comparison. Unfortunately, only a small number of dependent variables correlated with our demographic variables. Even in these cases, using the covariate failed to produce greater effect sizes in the planned comparisons, so demographic covariates are not reported in the results that follow.

Inquiry Behaviors

Table 1 shows the mean pre/post difference scores for the inquiry behaviors we coded in the video data, and Table 2 shows the results of the planned and post hoc comparisons conducted on these scores. Our planned comparisons, which employed analysis of variance (ANOVA) tests, revealed no effect of mediation on any of our measures; in other words, having an educator engage the family in an interactive tour of two exhibits without teaching them any skills did not have any effect on their subsequent inquiry behavior at a novel exhibit. However, in every planned comparison that showed an effect of inquiry, the two inquiry conditions outperformed the Exhibit Tour Control condition. (Effect sizes, measured by Cohen's d, ranged from 0.3 to 0.6.) Similarly, in every planned comparison that showed an effect of pedagogy, Juicy Question outperformed Hands-Off (with effect sizes ranging from 0.5 to 0.7).

In cases where differences were significant for inquiry, we conducted additional post hoc t-tests to determine which inquiry game was outperforming the Exhibit Tour Control condition. We also conducted post hoc t-tests when there were significant differences only for pedagogy, because such differences could be washed out when the two inquiry conditions were compared to control, resulting in no effect of inquiry. In every case, the post hoc tests found that Juicy Question significantly outperformed the Exhibit Tour condition, whereas Hands-Off outperformed Exhibit Tour in only one instance. Taken together, our planned comparisons and post hoc analyses indicate that the Juicy Question game, with its coaching in two target skills and collaborative pedagogy, more effectively promoted visitors' inquiry skills at a novel exhibit.

Juicy Question Game Improved Inquiry. Overall, Tables 1 and 2 show that the Juicy Question activity was effective at improving visitor-driven inquiry at interactive exhibits in a science museum. Because of the dominance of Juicy Question over both Hands-Off and Exhibit Tour, we focus our discussion on the performance of families in the Juicy Question condition.

Engagement. Compared to Exhibit Tour families, Juicy Question families increased their engagement with exhibits, spending significantly more time at the posttest than the pretest exhibit; families in all other conditions slightly decreased their time at exhibits from pre to post (see Figure 4). This increase for Juicy Question families could be due to

[6] To measure the impact of this discrepancy, we created a subsample of 152 families who were matched across the four conditions on number of boys in the family and reanalyzed the data in our 13 dependent measures. We found that 11 of the 13 measures were unaffected, the exceptions being (1) time at exhibit no longer showed an effect of inquiry and (2) linked PA string length, while continuing to show an effect of inquiry, showed no difference in the post hoc comparison between Juicy Question and Exhibit Tour. We have no way to determine whether these changes in the results are due to the demographic discrepancy or to the reduced number of families in the matched sample.

TABLE 1
Mean Pre/Post Difference Scores for Inquiry Behaviors Captured on Video

Inquiry Behaviors	Measures	Post Minus Pre Difference Scores by Condition: Mean (*SD*)			
		Juicy Question (Structured Inquiry)	Hands-Off (Spontaneous Inquiry)	Exhibit Tour (Control)	Pure Control (Control)
Engagement	Time at exhibit (minutes)	0.76 (3.9)	−0.19 (3.9)	−1.4 (4.9)	−1.0 (3.3)
Skill 1: Proposing actions	Number of PAs	2.8 (15)	0.26 (12)	−0.42 (12)	−1.1 (11)
	PAs per minute	0.26 (1.2)	0.07 (1.2)	0.34 (0.93)	0.17 (1.1)
	Duration of PAs (minutes)	0.32 (0.85)	0.22 (0.66)	−0.02 (0.69)	−0.05 (0.67)
	Level 2 PAs as percentage of PAs	−0.1 (14)	−1.8 (19)	−2.8 (14)	−7.1 (15)
Skill 2: Interpreting results	Number of IRs	16.3 (20.2)	4.20 (19.8)	−1.18 (22.1)	−1.54 (19.4)
	IRs per minute	1.3 (1.3)	0.43 (1.5)	0.43 (1.5)	0.26 (1.7)
	Duration of IRs (minutes)	0.77 (0.97)	0.56 (0.82)	0.27 (0.78)	0.06 (0.92)
	Level 2 IRs as percentage of IRs	11 (13)	5.1 (13)	5.3 (11)	2.6 (13)
Collaborative explanations	Consecutive IRs as percentage of IRs	10 (14)	4.4 (15)	0.0 (14)	0.0 (18)
	Consecutive IR string length	1.3 (2.2)	0.52 (2.2)	−0.32 (2.2)	−0.30 (2.3)
Coherent investigations	Linked PA string length	1.9 (4.0)	1.5 (3.6)	0.16 (3.3)	0.64 (2.9)
Other	Reads exhibit label	1.7 (1.5)	0.68 (1.3)	1.1 (1.1)	1.0 (1.4)

TABLE 2
Planned and Post Hoc Comparisons for Inquiry Behaviors Captured on Video

		Planned Comparisons (ANOVA): $F_{1,180}$ (effect size)			Post Hoc t-tests: t_{90} ($p < .05$)	
Inquiry Behaviors	Measures	Effect of Mediation (ET $>$ PC)	Effect of Inquiry (JQ $+$ HO $>$ ET)	Effect of Pedagogy (JQ $\neq$ HO)	JQ $\neq$ ET	HO $\neq$ ET
Engagement	Time at exhibit (minutes)	–	5.5 (0.4)	–	2.4	–
Skill 1: Proposing actions	Number of PAs	–	–	–	n/a	n/a
	PAs per minute	–	–	–	n/a	n/a
	Duration of PAs (minutes)	–	5.1 (0.4)	–	2.2	–
	Level 2 PAs as percentage of PAs	–	–	–	n/a	n/a
Skill 2: Interpreting results	Number of IRs	–	10.5 (0.6)	8.8 (0.6)	4.1	–
	IRs per minute	–	3.0 (0.3)	8.8 (0.6)	3.2	–
	Duration of IRs (minutes)	–	8.2 (0.5)	–	3.0	2.0
	Level 2 IRs as percentage of IRs	–	–	5.9 (0.5)	2.5	–
Collaborative explanations	Consecutive IRs as percentage of IRs	–	4.3 (0.4)	–	2.9	–
	Consecutive IR string length	—	9.9 (0.5)	–	3.6	–
Coherent investigations	Linked PA string length	–	6.1 (0.4)	–	2.3	–
Other	Reads exhibit label	–	–	13.7 (0.7)	2.1	–

Notes. The effects of mediation and of inquiry employed one-tailed tests; all other comparisons used two-tailed tests.
Effect sizes (Cohen's d) are shown in parentheses. Only results significant at the .05 p level are shown.
– = No significant differences.
ET = Exhibit Tour Control; PC = Pure Control; JQ = Juicy Question; HO = Hands-Off. n/a = Post hoc test not performed because planned comparisons revealed no differences.

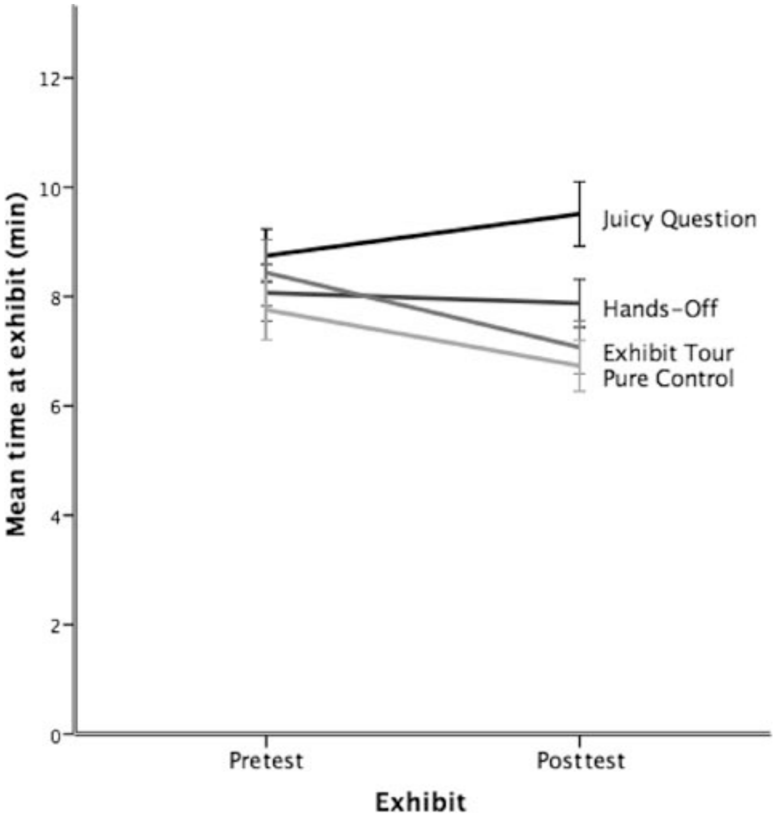

Figure 4. Pre/post change in mean time spent at the exhibits by families in each study condition. (In all graphs, error bars represent ± 1 *SE*.)

extended inquiry investigations, or it could be the result of following the time-consuming steps of the game. Subsequent analyses, reported below, support the former explanation.

Proposing Actions (Skill 1). Surprisingly, there was little effect of Juicy Question on the skill of PA. There was a nonsignificant increase in the number of PAs spoken by Juicy Question families, and no change in the percentage of PAs that were at level 2 (PA2s). However, the duration of PAs did increase significantly over control. For an example illustrating such an increase in duration, take an 11-year-old girl in the Juicy Question condition who gave only one, brief PA at the pretest exhibit: "You could put like two of them [together]." At the posttest exhibit, her PAs were longer, such as: "Let's say Daddy takes the middle one and starts swinging it back and forth. And I take this [end] one and start swinging it back and forth. I wonder what will happen to the ones in the middle, between Daddy's and my pendulum."

Interpreting Results (Skill 2). Juicy Question had a substantial impact on the interpretations families made about the phenomena they were exploring. Families in that condition made significantly more IR utterances than families in the Exhibit Tour condition. While this result might have simply reflected their longer times at the posttest exhibit, Juicy Question families also made significantly more interpretations per minute than control families, indicating that their inquiry actually became more lively (see Figure 5). Furthermore, their interpretations had longer durations on average and were more likely to be

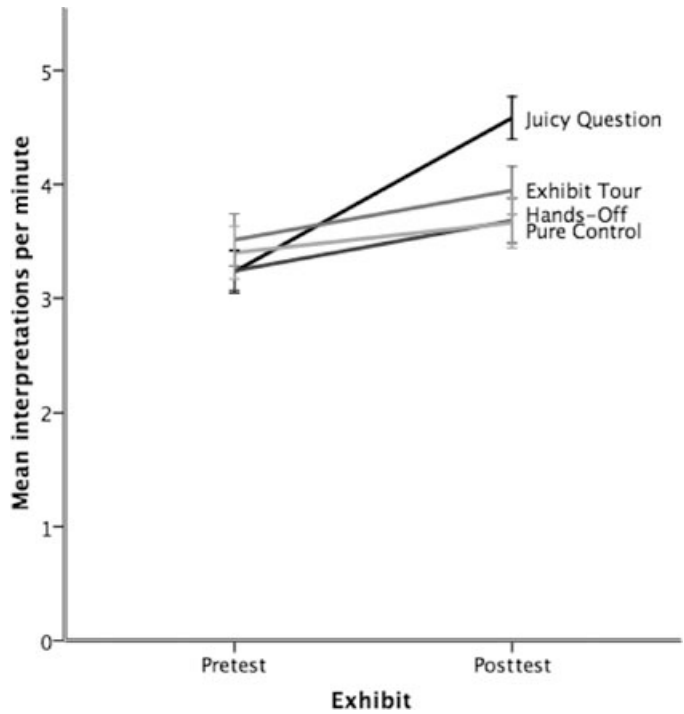

Figure 5. Pre/post change in mean number of IR utterances made per minute by families in each study condition.

at level 2 than those made by families in the Exhibit Tour condition, suggesting that the quality of this inquiry skill improved. (Hands-Off families also outperformed Exhibit Tour families on duration of IRs.) For example, a 13-year-old girl who made simple observations at the pretest exhibit such as, "Look at how everything is grouping together" (IR1) gave more explanatory interpretations, such as "Maybe the reason why it's stopping so abruptly is because it's in the middle" (IR2), when playing Juicy Question at the posttest exhibit.

Collaborative Explanations. Juicy Question also fostered more consecutive interpretations, in which family members respond to each other's explanations. Normalizing for number of IRs made by the families, Figure 6 shows a larger increase in the fraction of Juicy Question families' IRs that were consecutive. Moreover, the length of the longest string of consecutive IRs increased significantly more than in Exhibit Tour, suggesting that Juicy Question families had more vigorous, lengthy discussions about the results of their experiments. For example, the following conversation was generated by a Juicy Question family who increased their consecutive IRs substantially in the posttest. After conducting four investigations at the posttest exhibit, the family stopped to state their discoveries:

Boy: OK, so what did we find out?
Dad: What did we discover?
Mom: Yes, what did we discover?
Boy: I found that mine, they all started going in opposite directions. It looked like they were trying to make a wave, like keep going and trying to stop at the same time. That created a pretty neat effect.
Mom: But it ended up being the same energy.

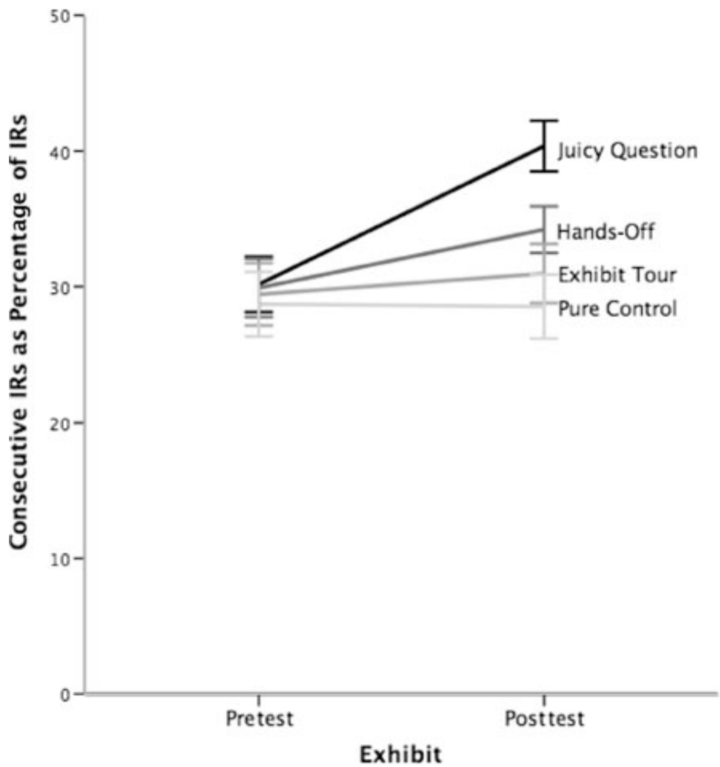

Figure 6. Pre/post change in mean percentage of IR utterances that were consecutive IRs, for each condition.

Boy:	Right.
Mom:	Eventually the energy synced up–
Boy:	It made a wave.
Mom:	–into the same wave–
Girl:	Right.
Mom:	–and then eventually into a parallel, linear thing.
Boy:	But it looked cool from it going, from everything going like this [in and out] to a wave.
Girl:	Yeah and like how mine got in sync, but after a while, it was like a little bit, because, Mommy, Daddy and I went different speeds, and somehow, like in less than half a minute, we all got in sync.
Mom:	Yeah, so matter what we did, eventually they all tried to work themselves in sync, and then the least resistance is the wave and then the linear wave, right?
Girl:	Right.
Boy:	Yup.
Mom:	Because as the energy gets expended. . . [End of mother's utterance]

In this example, mom and her two children worked together to understand the results of their experiments, illustrating deeper engagement with the inquiry process through collaborative building of explanations.

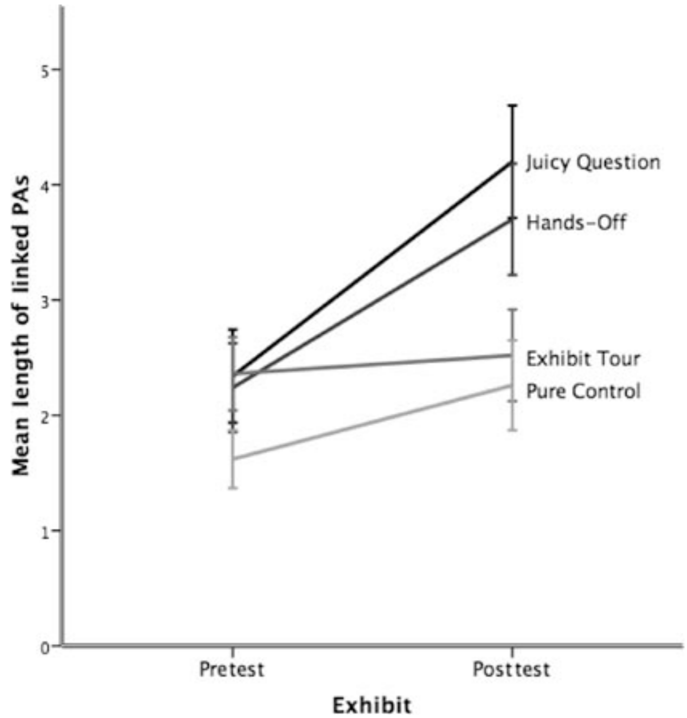

Figure 7. Pre/post change in mean string length of linked PAs (number of PAs that are linked), an indicator of coherent investigations, by families in each study condition.

Coherent Investigations. Juicy Question also had a significant impact on linked PAs, a measure of the topicwise coherence of families' investigations. As shown in Figure 7, multiple investigations, each started by a PA utterance, were more frequently connected by common themes in the Juicy Question condition than in the Exhibit Tour condition.

For example, one family conducted five linked investigations (for a linked PA score of 4) in pursuit of the following question at the posttest exhibit: Do you get a bigger wave by pushing one pendulum a *small* amount or a *large* amount?

Girl's PA:	If you swing it [a pendulum at one end] higher, will it [make a bigger wave]?
[Experiment 1:	They watch the pendulums swinging.]
Boy's PA:	Do it really high, as high as you can do. One, two three, and let go.
[Experiment 2:	They try a large initial push and then try several different things.]
Mom's PA:	Maybe we don't have to do it as hard as we can. Maybe the slower it [the initial push] is, the stronger the wave propulsion.
[Experiment 3:	They try it with a small initial push.]
Mom's IR:	It looks like it [the wave] is going faster.
[Experiment 4:	They try it again with a small initial push.]
Girl's IR:	There's probably like a medium amount of force, that.
Boy's PA:	I'm just gonna do a tiny bit and see.
[Experiment 5:	They try it with an even smaller initial push.]

The linked PA result shows that Juicy Question families became more deeply interested in a single question, drilling into it with longer chains of linked experiments. (The Hands-Off condition marginally outperformed the Exhibit Tour condition on this measure.)

In summary, various measures in the video data provided evidence that the Juicy Question game improved families' inquiry behaviors, particularly their ability to make generalizable or explanatory interpretations of the exhibit in a collaborative manner and to conduct multiple investigations on a common topic.

Family Members' Reflections on Their Experience

Immediately after the posttest exhibit, we interviewed one adult and one child from each family. We also phoned both participants 3 weeks later for a follow-up interview.

Posttest Ratings and Interview. Visitors in all conditions reported enjoying the experience. Five-point Likert scale questions asked adult and child visitors to rate whether they "had fun," "thought it was interesting," and felt it "took too long" (1 = never and 5 = always). Family members in all four conditions answered similarly, with the average responses in these three items ranging from 4.1 to 4.5 (with "take too long" reverse scored). While this is not a rigorously validated scale, we used it as a simple check that the inquiry games did not feel onerous to visitors or detract from the exhibit experience, an important characteristic of successful interventions in self-directed learning environments. The only statement that revealed a significant difference across groups was "I felt like I learned something," with adult visitors in the Exhibit Tour Control condition agreeing more strongly than adults in the Pure Control condition (planned ANOVA for effect of mediation: $F_{1,196} = 8.7$, $p < .01$).

Benefits and Drawbacks of the Mediated Programs. We asked visitors in the three mediated conditions to tell us what they liked most and least about their experience, and we coded their responses using an emergent set of codes that were applied by two coders. Fisher Exact tests were used to make planned comparisons on the effects of inquiry and pedagogy. Table 3 shows that families' characterizations varied across conditions.

Visitors in the Juicy Question condition more often mentioned that the game helped them think, focus, and collaborate. For example, one visitor in that condition said, "it gave us all a chance to interact and share ideas and stimulate thought." Another responded, "the kids came up with a hypothesis and tested it and really learned something, instead of just playing. A lot of times they learn from just playing, but this forced them to think about a scientific hypothesis."

Hands-Off visitors felt the game helped them collaborate by taking turns: "Everybody got to have their turns at discovering or trying things." Another Hands-Off visitor explained, "you would try stuff that I hadn't thought of. Someone else in the group would think of it and we'd try it and I'd see and think, 'Wow, that was cool.'"

Finally, the Exhibit Tour visitors enjoyed learning about how the exhibits were developed, hearing about the science content, and spending time with a good teacher. One Exhibit Tour visitor said of the staff educator, "I liked how she explained the initial idea [of the exhibit], the problems, then how they fixed them. That was interesting." Another liked it "when we did the air exhibit, on that exhibit when she talked about making [the ball] spin and how the faster air is on the top and that made it float." The Exhibit Tour visitors also made significantly fewer negative comments about their experience than visitors in the two inquiry conditions.

In terms of the negative, visitors in Juicy Question and Hands-Off mentioned several problems more often than visitors in the Exhibit Tour: The experience felt more rigid ("sometimes it seemed a little forced"); it was difficult to get everyone to participate ("the fact that when you don't say hands off, and just hop in to try something, it's not respectful,

TABLE 3
Visitors' Responses to the Mediated Conditions in the Posttest Interview

	Children			Fisher *p*		Adults			Fisher *p*	
	Juicy Question	Hands-Off	Exhibit Tour	Inquiry	Pedagogy	Juicy Question	Hands-Off	Exhibit Tour	Inquiry	Pedagogy
Liked most										
Helped us think	**18%**	0%	0%	**.05**	**.01**	**52%**	16%	4%	**.01**	**.01**
Helped us focus our activity	**58%**	12%	2%	**.01**	**.01**	**58%**	24%	4%	**.01**	**.01**
Helped us collaborate	**18%**	**34%**	2%	**.01**	–	**24%**	**36%**	2%	**.01**	–
Helped us take turns	0%	**56%**	2%	**.01**	**.01**	0%	**56%**	0%	**.01**	**.01**
Enjoyed learning about how the exhibits were developed	0%	0%	**36%**	**.01**	–	0%	0%	**56%**	**.01**	–
Enjoyed hearing about the science content	0%	0%	**36%**	**.01**	–	0%	0%	**42%**	**.01**	–
Enjoyed spending time with a good teacher	0%	0%	**14%**	**.01**	–	2%	0%	**30%**	**.01**	–
Liked least										
Experience felt rigid	20%	10%	4%	–	–	**26%**	**30%**	4%	**.01**	–
Difficult to get every family member to participate	**8%**	**10%**	0%	**.05**	–	**8%**	**20%**	0%	**.01**	–
Hard to maintain the game	4%	4%	0%	–	–	**14%**	**26%**	0%	**.01**	–
Difficult to decide upon a question to investigate	**34%**	2%	0%	**.01**	**.01**	**38%**	2%	0%	**.01**	**.01**
Felt tested at times	12%	2%	2%	–	–	**24%**	0%	6%	–	**.01**
Difficult to stop someone who had called "Hands-Off!"	6%	**44%**	4%	**.01**	**.01**	6%	16%	4%	–	–
No negative comments	30%	32%	**76%**	**.01**	–	12%	24%	**62%**	**.01**	–

Notes. Response categories are not mutually exclusive; percentages do not sum to 100%. Bold indicates significantly higher percentage of responses.
– = Nonsignificant result.

and it doesn't feel good when others do that to you"); and it was sometimes hard to maintain the game ("my mind was moving ahead to 'let's do that' rather than stopping to say hands off"). In Juicy Question, visitors felt it was sometimes difficult to decide upon a question to investigate, and at times they felt tested. One visitor mentioned the difficulty of generating a good question, stating, "I don't really think I got the wave [exhibit], so it made it hard to come up with a question because I didn't know what we were doing." The most common complaint among children in the Hands-Off condition was that it was often difficult to get a turn, once someone else had called "hands off!" at the exhibit. One protested, "We had to wait sometimes a lot."

Using the Game in the Future. We also asked visitors in the two inquiry conditions whether they thought they would use the game at exhibits in the museum, or indeed anywhere else in their lives. We found a significant difference, with 53% of Juicy Question visitors predicting that they would use the game, but only 30% of Hands-Off visitors making the same prediction (adults $\chi 2 = 11.0$, $p < .01$; children $\chi 2 = 7.9$, $p < .05$). Visitors in the Juicy Question condition often mentioned how the game helped them use exhibits to learn together:

> Because it opens up the communication between parents and children, it's more about communication than about just touching the exhibits. When we first came in, I don't think we were as open to looking at questions, rather than just doing the hands-on things. I can see the Juicy Question thing coming in handy with other adults and with children. This really helps you interact with your group, as well as just doing the hands-on things. I mean, out there, and on the way home, we can ask, "What do you think? What's going on?" (Adult)

> Because the Juicy Question game really helps; because without it, we wouldn't ask as many questions. I'm more likely to use it because it gets you more into the exhibit than you probably would. (Child)

Hands-Off visitors tended to focus on how the game established fair turn taking and inclusiveness in the learning process:

> I think it just makes it fair, so that everybody's idea can be put to use and everyone can see their own ideas used, it would make it easier for everyone, especially people who aren't forceful. (Adult)

> I would probably use it because there's going to be a lot of people here and you can just call "hands off" and get a chance to play with it. (Child)

Follow-Up Interview. While we would have preferred to observe the families during the rest of their visit and their lives beyond it, we failed to find a method for capturing spontaneous behavior without violating human protection consent requirements. Instead, we called the families 3 weeks after they had participated in the study and asked them to self-report whether there was anything they had learned during the study that they had used subsequently. Table 4 shows the responses of children and adults from each condition and the results of Fisher Exact tests for the effect of inquiry.

As we had expected, visitors in the two inquiry game conditions more often reported they had applied the two taught skills (PA or IR) at new exhibits and even outside the museum.

TABLE 4
Visitors' Responses in Follow-Up Interview

		Children					Adults				
Question	Response	Juicy Question	Hands-Off	Exhibit Tour	Pure Control	Fisher *p*	Juicy Question	Hands-Off	Exhibit Tour	Pure Control	Fisher *p*
Used anything the teacher said at other exhibits in the museum, after leaving the study room that day?	Used specific skills taught (PA or IR)	**34%**	**20%**	0%	n/a	**.01**	**19%**	**15%**	3%	n/a	**.05**
	Exhibit topic was relevant	3%	0%	**33%**	n/a	**.01**	0%	0%	**23%**	n/a	**.01**
	Viewed exhibits differently	0%	0%	0%	n/a	–	7%	0%	**18%**	n/a	**.05**
	Hard to implement	3%	16%	0%	n/a	–	**21%**	**38%**	0%	n/a	**.01**
Used anything the teacher said in your life after you left the museum that day?	We used the skills we learned	**17%**	**8%**	0%	n/a	**.05**	**12%**	**12%**	0%	n/a	**.05**
	The exhibit topics were relevant	0%	0%	**30%**	n/a	**.01**	0%	0%	**8%**	n/a	**.05**
	Didn't come up outside museum	**38%**	**60%**	15%	n/a	**.01**	34%	32%	28%	n/a	–
	We talked about the experience	3%	0%	4%	n/a	–	7%	6%	**31%**	n/a	**.01**
Remembered something about the exhibits?[a]	Remembered pretest exhibit	85%	91%	92%	96%	–	84%	75%	79%	84%	–
	Remembered second exhibit	100%	100%	100%	100%	–	92%	94%	94%	92%	–
	Remembered third exhibit	73%	70%	88%	74%	–	61%	66%	74%	61%	–
	Remembered posttest exhibit	73%	74%	68%	74%	–	53%	72%	50%	53%	–

Notes. Fisher Exact tests were employed to assess the effect of inquiry (JQ + HO > ET). Bold indicates significantly higher percentage of responses.
– = Nonsignificant result.
n/a = Not applicable.
[a]Data for this question were coded by a single coder, with 10% recoded by a second coder. Interrater agreement was 95%.

For example, an adult who had played Juicy Question said,

> We were playing with a lens exhibit [after the study] . . . and we were asking "I wonder if" questions, and then asking all the ways we could think about the exhibit. We found ourselves using the strategies again. It just became incorporated in how we talked at all the exhibits. People were looking at us a lot as they were waiting for the exhibit behind us, but it made it fun.

Outside the museum, a child from the Juicy Question condition recalled using the game on the drive home: "When we were going back home on the Bay Bridge after the museum, we were asking about the cranes out there, how did they stay out there." An adult who had learned Hands-Off recalled, "It was with that stroboscopic guitar [exhibit]. My daughter said something like 'I want to try this.' Whatever the catch phrase was, she said it. We all stopped and we listened to her try what she wanted to try."

Also as expected, the Exhibit Tour respondents more often said they applied the science content they had learned to related situations. For example, one of the exhibits in the tour demonstrates the Bernoulli effect and its relationship to air pressure. One family remembered discussing this topic again while visiting another museum: "The girls remembered at the Tech Museum the next day about the air pressure and why the balls floated." Another family talked about airplanes on the drive home: "I explained that to him in the car on the way home. He asked me how it makes planes fly and I explained the pressure difference. I explained the whole thing with reference to the balls in the air at the exhibit."

In short, the interview data suggest that

(i) visitors in all three mediated conditions enjoyed the experience they received, but that what they liked and disliked varied. Compared to controls, families who played the inquiry games particularly appreciated the way the game helped them to collaborate and (for Juicy Question) to think more deeply about the exhibits; at the same time, some families disliked the rigid participation rules of the games and (for Juicy Question) the challenge of finding an appropriate question to investigate.

(ii) while more Juicy Question families than Hands-Off families anticipated that they would spontaneously use the game after the study, there were no differences in number of self-reports of actual use 3 weeks later. An average of 21% of participating adults and 32% of participating children who received some kind of staff mediation reported spontaneously using something they had learned in a subsequent situation; most frequently, families in the inquiry game conditions applied the skills whereas Exhibit Tour families applied the exhibit content.

There were no significant differences across conditions in visitors' overall memories of the experience, as indicated by the number of exhibits they described having used. Nor were there differences in the patterns of responses by adults and children, suggesting that the families experienced the inquiry games coherently as a family group.

Science Content Correctness

This study was designed to focus on generalizable inquiry skills, rather than any particular scientific content. At the same time, we were interested to see whether the data suggested that families who learned to play the inquiry games might be learning more scientific content than the controls. Any such learning would be specific to each exhibit, because the exhibits were not thematically linked in terms of their science content. With families already burdened by the large existing battery of assessments, we could not include an

TABLE 5
Codable IR2s Made by Families in the Posttest

Condition	Codable IR2s in the Posttest	Mean Codable IR2s per Group
Juicy Question	415	8.3
Hands-Off	198	4.0
Exhibit Tour	195	3.9
Pure Control	167	3.3

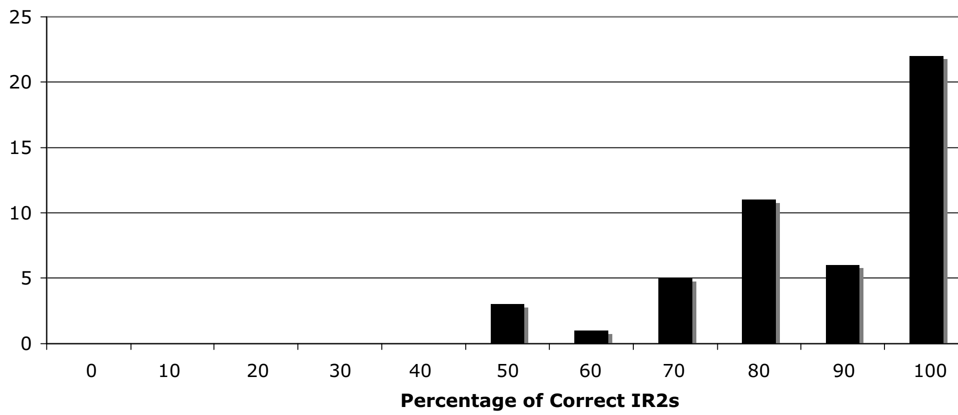

Figure 8. Number of families in the Juicy Question condition making correct interpretations (correct IR2s) as a percentage of their total interpretations (codable IR2s).

explicit content assessment, but we did code all families' high-level instances of IR (IR2s) for any statements that could be coded as *correct* or *incorrect* in relation to scientific canon. Unfortunately, families in all conditions made too few IR2s in the pretest (less than three on average) to warrant scoring for percentage correct, so only posttest IR2s were coded and scored. Furthermore, families in the Hands-Off and control conditions did not make enough IR2s, even in the posttest, to justify any kind of comparison (see Table 5). Consequently, our analysis examines only the IR2s made by families in the Juicy Question condition.

Figure 8 shows that nearly all of the Juicy Question families who made abstract or explanatory interpretations (IR2s) in the posttest were scientifically correct in the great majority of their assertions. In fact, almost half of the families (22 of 48) made canonically correct interpretations every time. Recall that families in this condition significantly increased their IR2s as a result of instruction. Taken together, these results indicate that Juicy Question families displayed mostly correct, higher level interpretations after instruction.

DISCUSSION

The broad pattern of results from our experiment showed that playing the Juicy Question game significantly improved families' inquiry processes at hands-on exhibits in a science museum. Specifically, Juicy Question increased the quantity and quality of families' interpretations, fostered more collaborative explanations, and promoted more coherent investigations, than controls. A majority of interview participants who learned the Juicy Question game felt it helped their family think and focus, but about a quarter of them also felt the game was rigid. Three weeks after the experience, about one fifth to one third of visitors said they had used it at exhibits in the museum or in their lives outside.

Of the three hypotheses put forth at the beginning of the study, two were supported and one was disconfirmed:

Hypothesis 1: Inquiry games will improve inquiry behaviors. On the basis of the strength of the design principles described earlier and the iterative development process we undertook, we were not surprised that the inquiry games were more successful than the Exhibit Tour. Specifically, Juicy Question outperformed control on many key inquiry behaviors, and even one behavior that is central to good science but was never explicitly part of either game—coherent investigations, as measured by linked PAs. The only inquiry behavior that did not show improvement was the number of PAs, either total or per minute. This is a curious result, since Juicy Question families were explicitly asked to generate PAs. Only the duration of their PAs increased.

Hypothesis 2: Mediation will enhance the experience. We were surprised that the mediated control condition, in which the staff educator gave an interactive Exhibit Tour, did not outperform the Pure Control condition on any measure. We attribute this to two major factors: By design, the educator was not with the family during the posttest, so any influence she may have had in heightening their interest or extending their engagement apparently had no residual effect when the family moved to the posttest exhibit. Second, it may well be that the Exhibit Tour group learned more accurate scientific content about the specific exhibits in the tour, but this did not transfer to the novel exhibit used in the posttest. Thus, the educator's Exhibit Tour may have had various localized influences on families' learning, but it had no effect on their transferable inquiry skills at a new exhibit.

Hypothesis 3: Pedagogy will affect inquiry. We predicted that Juicy Question might support deeper inquiry than Hands-Off, while Hands-Off might be easier for families to use and adopt. The results showed that that Juicy Question did support deeper inquiry; in fact, it was more effective even than Hands-Off at improving the frequency and sophistication of families' interpretations. However, Hands-Off was not necessarily easier to adopt, as both games were readily learned by families in the available time. In fact, as described below, some families seemed to have rejected the Hands-Off game as unnecessary.

Understanding Juicy Question's Success

To gain some qualitative sense of why Juicy Question outperformed Hands-Off on several measures, we reviewed video and interview data from a total sample of 22 high-improving and low-improving families in both conditions. While this review was not rigorous, it surfaced four characteristics of Juicy Question's activity structure (absent from the Hands-Off game) that seemed to have contributed to families' deeper inquiry behaviors.

Everyone Is Asked to Participate. In Juicy Question, all family members are asked to generate a question (PA) before the experiment and voice discoveries (IR) at the end. Hands-Off makes no such request of the whole group; individuals are free to remain silent throughout the interaction with the exhibit. When everyone is asked to practice the two inquiry skills, these skills may be employed more often, role-modeled more frequently by adults, and integrated into the group's interactions with the exhibit.

The Activity Requires Collaboration. In Juicy Question, the family must negotiate when choosing a question to pursue. This supports their collaboration skills and may heighten each person's investment in the outcome of the experiment. By requiring discussion of which question to pursue, the game may even improve the quality of the experimental designs, leading to better experiments that evoke more interpretations. In contrast, Hands-Off families simply yield the floor to the person calling "hands off."

The Activity Requires Family Members to Articulate Their Interpretations. The Juicy Question game includes an explicit phase in which family members are asked to state their interpretations, reinforcing that inquiry skill (Quintana et al., 2004). Hands-Off also teaches the skill but does not require anyone to use it. In particular, there is less need to call "hands off" for a discovery than for a plan, because stating a discovery rarely requires control of the exhibit. This disparity between the two games may be partly responsible for the difference in the frequency with which visitors interpreted results in the two conditions.

Collaboration May Add More Value Than Turn Taking. The incentive for playing the Juicy Question game, beyond merely pleasing the museum educator, is improved collaboration: The game offers families a process for working together to make their experiments more interesting and their interpretations more engaging. In Hands-Off, the main incentive is improved turn taking; calling "hands off" allows family members to gain temporary control of the exhibit. However, if some families feel they do not need help taking turns, they may see no benefit to playing the Hands-Off game and thus fail to adopt it as willingly as families in the Juicy Question condition. In our review of the videos and interviews, several of the low-improving Hands-Off families actually joked derisively about the game. One man who learned Hands-Off commented in the posttest interview, "I kinda kept making fun of it a little bit... the cards seemed like a silly extra added element." In the follow-up interview, another man from the Hands-Off condition said he did not think the museum should offer the game to visitors, because turn taking at exhibits is not a problem: "Everybody [in the museum] understood, on all the exhibits that we tried, to take turns, so it went smoothly. And even when little kids tried to interfere, their parents kept them in check. There just really wasn't a need." It is possible that the benefit for playing Hands-Off—improved turn taking—was not great enough to persuade some families to adopt it.

These characteristics of Juicy Question which point to its success map well onto the design principles for inquiry curricula mentioned earlier, namely modeling, articulation, reflection, and collaboration. In creating the inquiry games, we knew that Juicy Question emphasized these principles to a greater degree, but thought that Hands-Off better supported families' own learning agendas by allowing family members to choose when to participate. In the end, the success of an inquiry game seemed to hinge more on its emphasis of inquiry pedagogy than its capacity to support families' varied learning agendas.

In summary, it seems that teaching the two inquiry skills in a game was necessary but insufficient for changing families' behaviors; the activity structure of the Juicy Question game was critical for improving inquiry. Apparently, a more formal structure that encourages group collaboration and participation is necessary for enacting real change in families' inquiry behaviors in an informal setting.

Limitations of the Study

This study has several limitations that may narrow the applicability of the results. First, even though the families were left alone to use the last (posttest) exhibit, before leaving the educator did explicitly ask families in the two inquiry conditions to play the inquiry game that they had learned. One could argue that asking these visitors to play Juicy Question or Hands-Off at the posttest exhibit somewhat undermines our ability to make strict comparisons among the conditions and requires that we report our results with the caveat, "when prompted to play the games." We chose to do this because the alternative—asking visitors to use a final exhibit "as they normally would" after we had given them game cards and spent 20 minutes teaching them an exhibit game—seemed disingenuous; also it would have led to a confounding, because we would not have known whether families freely chose to play the game or succumbed to an obvious pressure to please the educator. By cuing everyone to use it, we created a best-case scenario for determining how well visitors could play the games at the posttest exhibit. More importantly, even in ideal circumstances, simply telling someone to engage in a particular learning process does not mean they will have the motivation or ability to comply. Barbara White, an experienced inquiry researcher and project advisor, commented that "We teach inquiry to kids in schools for 13 weeks, and often they still don't do it in a new situation when we ask" (B. Y. White, personal communication, February 14, 2005). Similarly, our experience during the early development of the inquiry games made it clear that giving families a broad directive to play the game was frequently ignored, especially if the families did not find the game memorable or enjoyable.

A second limitation is that we coached and assessed inquiry behaviors at only one type of science museum exhibit, which we characterized as interactive, open-ended, multioption, and multiuser.[7] The results of the study may be less applicable to so-called "planned discovery" exhibits, which have fewer options to try (Hein, 1998; Humphrey & Gutwill, 2005) or to static display exhibits such as dioramas.

Finally, we conducted our study in a laboratory, rather than on the museum's exhibit floor. The experience was authentic insofar as the families were genuine museum visitors, and the exhibits were in the exact form in which they would normally be available on the museum floor. Still, our laboratory provided a best-case scenario, in that families had the exhibits to themselves, could spend as long as they wanted, could hear each other clearly, and endured few distractions while using each exhibit. How well would the Juicy Question game function on the buzzing museum floor, where a family may not have complete control over the exhibit, and others may be waiting in line to use it? Further research is needed to answer this question.

Implications for Family Inquiry

Despite these limitations, the results of this study yield implications for our understanding of family inquiry beyond the museum context. First, this study suggests that offering parents (or care-givers) a structured, coinvestigative role in exploring phenomena may significantly enhance families' inquiry. Parents are natural guides for their children, helping them make sense of novel informal science settings, but parents' explanations of phenomena are often brief and fragmented (Crowley & Galco, 2001). Moreover, parents may not naturally take the role of coinvestigator, instead teaching their children didactically (Shine & Acosta, 1999) or delegating to their children simple tasks while practicing higher level ones themselves such as making inferences (Gleason & Schauble, 2000). In the Juicy Question game, we asked parents to facilitate the experience—playing to their natural tendency to guide their children—but we also challenged them to pursue questions to which no one knows the answer. This latter request successfully encouraged both parents and children to actively investigate and reflect on the exhibit, as demonstrated by their improvement in the skill of

[7] In a technical sense, the results could be deemed even more limited, applying only to our posttest exhibit, Making Waves. However, we suggest that improvement at the posttest exhibit in the Juicy Question condition would have required successful application of that game at the two "treatment" exhibits, Floating Objects and Unstable Table. Consequently, our finding that the Juicy Question game facilitates family inquiry arguably pertains to three distinct exhibits.

interpreting results. It seems plausible that parents may embrace opportunities to take on coinvestigative roles, thereby enhancing family inquiry in informal settings more generally.

Another implication from this study is that asking family members to articulate their thoughts may be critical to deepening their inquiry. The collaborative structure of the Juicy Question game—brainstorming, pursuing, and then reflecting on a question chosen by the group—rests upon the simple request that everyone explicitly state their questions and discoveries. As discussed above, the difference in performance of families in the Juicy Question and Hands-Off conditions may be due to the fact that the former game required such articulation whereas the latter did not. Research in classrooms has similarly shown that asking students to state their thoughts aloud results in improved problem-solving performance (Fletcher, 1985). And in art museums, the VTS approach has been shown to improve field trip students' critical thinking skills, in part by asking them to articulate their observations about an art piece under discussion (Housen, 2001). Perhaps a key component of family inquiry, indeed even group inquiry in general, is that members verbalize their ideas.

Future Directions

In light of this research, we plan to bring the Juicy Question game to families on the actual museum floor. A new set of questions will need to be addressed: What kinds of visitors would most want to learn and use the game? Would regular family visitors be interested, or should we begin with museum members and homeschoolers? How and when in their visit should families be approached by our educators? How should we tailor the game to any particular audience we engage? We hope to find answers to these and other questions.

Another group that is an obvious audience for Juicy Question is school students and chaperones on field trips. In fact, a second phase of this project, already underway, studies field trip groups using the same experimental design as reported here. Preliminary results suggest that the Juicy Question condition is again highly successful at improving students' inquiry processes. If these initial results hold, our upcoming project to bring Juicy Question to the museum floor will include school and camp field trip groups.

The authors gratefully acknowledge the creative and analytical contributions of all members of the GIVE team, particularly Mark Boccuzzi, Beth Gardner, Malia Jackson, Adam Klinger, Nerissa Kuebrich, Suzy Loper, Kathleen McLean, Lisa Sindorf, and Erin Wilson. We also want to thank Linda Deacon, John Frederiksen, Michael Ranney, Barbara White, Laura Wise, and the advisors to the GIVE project. Finally, we thank the editor and three anonymous reviewers for their insightful comments on a previous version of this article. We are grateful for the generous financial support of the National Science Foundation.

REFERENCES

Adamson, B. (2008). The relationship between exhibition environments and visitors' behaviors. Paper presented at the Visitor Studies Association, Houston, TX.

Adey, P. & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention programme in the high-school science curriculum. Cognition and Instruction, 11(1), 1–29.

Allen, S. (1997). Using scientific inquiry activities in exhibit explanations. Science Education (Informal Science Education—Special Issue), 81(6), 715–734.

Allen, S. (2002). Looking for learning in visitor talk: A methodological exploration. In G. Leinhardt, K. Crowley, & K. Knutson (Eds.), Learning conversations in museums (pp. 259–303). Mahwah, NJ: Erlbaum.

Allen, S. (2004a). Designs for learning: Studying science museums exhibits that do more than entertain. Science Education, 88 (Suppl. 1), S17–S33.

Allen, S. (2004b). Finding significance. San Francisco: Exploratorium.
Allen, S. (2004c). To partition or not to partition: A study of the impact of walls on visitor behavior at an exhibit cluster. Paper presented at the Visitor Studies Association, Albuquerque, NM.
Allen, S., & Gutwill, J. P. (2009). Creating a program to deepen family inquiry at interactive science exhibits. Curator, 52(3), 289–306.
Ansbacher, T. (1999). Experience, inquiry and making meaning. Exhibitionist, 18(2), 22–26.
Ash, D. (1999). The process skills of inquiry. In L. Rankin, D. Ash, & B. Kluger-Bell (Eds.), Inquiry: Thoughts, views and strategies for the K-5 classroom. Washington, DC: National Science Foundation.
Bakeman, R. & Gottman, J. M. (1997). Observing interaction: An introduction to sequential analysis (2nd ed.). Cambridge, England: Cambridge University Press.
Bell, P., Lewenstein, B., Shouse, A., & Feder, M. (Eds.). (2009). Learning science in informal environments: People, places, and pursuits. Washington, DC: National Academies Press.
Borun, M. J., Dritsas, J. I., Johnson, N. E., Peter, K. F., Fadigan, K., Jangaard, A., et al. (1998). Family learning in museums: The PISEC perspective. Washington, DC: Association of Science-Technology Centers.
Bransford, J., Brown, A., & Cocking, R. R. (Eds.). (2003). How people learn: Brain, mind, experience and school. Washington, DC: National Academy Press.
Brown, A. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. In H. W. Reese (Ed.), Advances in child development and behavior (Vol. 10, pp. 103–152). New York: Academic Press.
Champagne, A. B., Kouba, V. L., & Hurley, M. (2000). Assessing inquiry. In J. Minstrell & E. H. v. Zee (Eds.), Inquiring into inquiry learning and teaching in science (pp. 447–470). Washington, DC: American Association for the Advancement of Science.
Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145–182.
Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Science Education, 86(2), 175–218.
Collins, A., Brown, J. S., & Newman, S. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), Knowing, learning, and instruction (pp. 453–494). Hillsdale, NJ: Erlbaum.
Crowley, K. & Galco, J. (2001). Everyday activity and the development of scientific thinking. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), Designing for science: Implications from everyday, classroom, and professional settings. Mahwah, NJ: Erlbaum.
Crowley, K. & Jacobs, M. (2002). Building islands of expertise in everyday family activity. In G. Leinhardt, K. Crowley, & K. Knutson (Eds.), Learning conversations in museums. Mahwah, NJ: Erlbaum.
Csikszentmihalyi, M. & Hermanson, K. (1995). Intrinsic motivation in museums: Why does one want to learn? In J. H. Falk & L. D. Dierking (Eds.), Public institutions for personal learning: Establishing a research agenda (pp. 67–77). Washington, DC: American Association of Museums.
Davis, J. (1993). Museum games. Teaching Thinking and Problem Solving, 15(2), 1–6.
Diamond, J. (1986). The behavior of family groups in science museums. Curator, 29(2), 139–154.
Doering, Z. (1999). Strangers, guests, or clients? Visitor experiences in museums. Curator, 42(2), 74–87.
Ellenbogen, K. (2002). Museums in family life: An ethnographic case study. In G. Leinhardt, K. Crowley, & K. Knutson (Eds.), Learning conversations in museums. Mahwah, NJ: Erlbaum.
Falk, J. H. (2009). Identity and the museum visitor experience. Walnut Creek, CA: Left Coast Press.
Falk, J. H. & Dierking, L. (2000). Learning from museums: Visitor experiences and the making of meaning. New York: Alta Mira Press.
Falk, J. H., Moussouri, T., & Coulson, D. (1998). The effect of visitors' agendas on museum learning. Curator, 41(2), 106–120.
Flavell, J. H. (1973). Metacognitive aspects of problem-solving. In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, NJ: Erlbaum.
Fleiss, J. L., Levin, B., & Paik, M. C. (2004). The measurement of interrater agreement. In Statistical methods for rates and proportions (3rd ed., pp. 598–626). Hoboken, NJ: Wiley.
Fletcher, B. (1985). Group and individual learning of junior high school children on a micro-computer-based task. Educational Review, 37, 252–261.
Gleason, M. E. & Schauble, L. (2000). Parents' assistance of their children's scientific reasoning. Cognition and Instruction, 17(4), 343–378.
Gutwill, J. P. (2005). Observing APE. In T. Humphrey & J. Gutwill (Eds.), Fostering active prolonged engagement: The art of creating APE exhibits. San Francisco: Left Coast Press.
Gutwill, J. P. (2008). Challenging a common assumption of hands-on exhibits: How counter-intuitive phenomena can undermine open-endedness. Journal of Museum Education, 33(2), 187–198.

Hein, G. E. (1998). Learning in the museum. New York: Routledge.
Hestenes, D. (1987). Toward a modeling theory of physics instruction. American Journal of Physics, 55(5), 440–454.
Hilke, D. D. (1985). The family as a learning system: An observational study of families in museums. In B. H. Butler & M. B. Sussman (Eds.), Museum visits and activities for family life enrichment (pp. 101–129). New York: Hayward Press.
Hollander, E. & Harvey, R. J. (2002). Generalizability theory analysis of item-level O*NET database ratings. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.
Housen, A. C. (2001). Aesthetic thought, critical thinking and transfer. Arts and Learning Research Journal, 18(1), 99–131.
Housen, A. C. & Yenawine, P. (2001). Basic VTS at a glance. New York: Visual Understanding in Education.
Humphrey, T. & Gutwill, J. P. (Eds.). (2005). Fostering active prolonged engagement: The art of creating APE exhibits. Walnut Creek, CA: Left Coast Press.
Kuhn, D., & Udell, W. (2003). The development of argument skills. Child Development, 75(5), 1245–1260.
Labudde, P., Reif, F. & Quinn, L. (1988). Facilitation of scientific concept learning by interpretation procedures and diagnosis. International Journal of Education, 10(1), 81–98.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159–174.
National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press.
National Research Council. (2000). Inquiry and the National Science Education Standards: A guide for teaching and learning. Washington, DC: National Academy Press.
National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: National Academies Press.
National Research Council. (2009). Learning science in informal environments: People, places, and pursuits. Washington, DC: National Academies Press.
Oppenheimer, F. (1968). A rationale for a science museum. Curator, 11(3), 206–209.
Oppenheimer, F. (1972). The Exploratorium: A playful museum combines perception and art in science education. American Journal of Physics, 40(July), 978–984.
Palincsar, A. & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 1, 117–175.
Perry, D. (1993). Designing exhibits that motivate. In M. B. Patty McNamara, S. Grinell, & B. Serrell (Ed.), What research says about learning in science museums (pp. 25–29). Washington, DC: Association of Science-Technology Centers.
Piaget, J. (1978). Success and understanding. Cambridge, MA: Harvard University Press.
Piscitelli, B., & Weier, K. (2002). Learning with, through, and about art: The role of social interactions. In S. G. Paris (Ed.), Perspectives on object-centered learning in museums (pp. 121–151). Mahwah, NJ: Erlbaum.
Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A scaffolding design framework for software to support science inquiry. Journal of the Learning Sciences, 13(3), 337–386.
Randol, S. (2005). The nature of inquiry in science centers: Describing and assessing inquiry at exhibits. Unpublished Doctoral Dissertation, University of California, Berkeley.
Samarapungavan, A., Mantzicopoulos, P., & Patrick, H. (2008). Learning science through inquiry in kindergarten. Science Education, 92(5), 868–908.
Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), Liberal education in a knowledge society (pp. 67–98). Chicago: Open Court Publishing.
Semper, R. J. (1990). Science museums as environments for learning. Physics Today, 43(11), 50–56.
Shine, S., & Acosta, T. Y. (1999). The effect of the physical and social environment on parent-child interactions: A qualitative analysis of pretend play in a children's museum. In S. Reifel (Ed.), Play contexts revisited (pp. 123–142). New York: Greenwood.
Songer, N. (2004). Persistence of inquiry: Evidence of complex reasoning among inner city middle school students. Paper presented at the American Educational Research Association, San Diego, CA.
St. John, M., Carroll, B., Helms, J., & Smith, A. (2008). The PIE Institute project: Final evaluation report. San Francisco: Inverness Research Associates.
Tien, L. T., Rickey, D., & Stacey, A. (1999). The MORE thinking frame: Guiding students' thinking in the laboratory. Journal of College Science Teaching, 18(5), 318–324.
Tisdal, C. & Perry, D. (2003). Going APE at the Exploratorium: Interim summative evaluation report. Chicago: Selinda Research Associates.

Tobin, K., Tippins, D., & Gallard, A. J. (1994). Research on using laboratory instructions in science. In D. L. Gabel (Ed.), Handbook of research on science teaching and learning (pp. 45–93). New York: National Science Teachers Association.

von Glasersfeld, E. (1991). Cognition, construction of knowledge, and teaching. In M. Matthews (Ed.), History, philosophy and science teaching. Toronto, Ontario, Canada: Teacher's College Press.

Vygotsky, L. S. (1962). Thought and language. Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978). Mind and society: The development of higher mental processes. Cambridge, MA: Harvard University Press.

White, B. (1993). ThinkerTools: Causal models, conceptual change, and science education. Cognition and Instruction, 10(1), 1–100.

White, B. & Frederiksen, J. (1998). Inquiry, modeling and metacognition: Making science accessible to all students. Cognition and Instruction, 16(1), 3–118.

White, B. & Frederiksen, J. (2000). Metacognitive facilitation: An approach to making scientific inquiry accessible to all. In J. Minstrell & E. H. v. Zee (Eds.), Inquiring into inquiry learning and teaching in science (pp. 331–370). Washington, DC: American Association for the Advancement of Science.

White, R. T. (1988). Learning science. Oxford, England: Basil Blackwell.

Wood, D. (2001). Scaffolding, contingent tutoring and computer-supported learning. International Journal of Artificial Intelligence in Education, 12, 280–292.