Deepening Students' Scientific Inquiry Skills During a Science Museum Field Trip

Joshua P. Gutwill \(^a\) & Sue Allen \(^a\)\(^b\)

\(^a\) Exploratorium, San Francisco, California, USA
\(^b\) National Science Foundation, Arlington, Virginia, USA

Available online: 24 Jun 2011

To cite this article: Joshua P. Gutwill & Sue Allen (2012): Deepening Students’ Scientific Inquiry Skills During a Science Museum Field Trip, Journal of the Learning Sciences, 21:1, 130-181

To link to this article: http://dx.doi.org/10.1080/10508406.2011.555938

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages.
whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Field trips to science museums can provide students with educational experiences, particularly when museum programs emphasize scientific inquiry skill building over content knowledge acquisition. We describe the creation and study of 2 programs designed to significantly enhance students’ inquiry skills at any interactive science museum exhibit without the need for advanced preparation by teachers or chaperones. The programs, called Inquiry Games, utilized educational principles from the learning sciences and from visitor studies of museum field trips. A randomized experimental design compared 2 versions of the games to 2 control conditions. Results indicate that the groups that learned the Inquiry Games significantly outperformed the control groups in the duration and quality of several inquiry skills when using a novel exhibit, with effect sizes ranging from 0.3σ to 0.8σ. The highest gains came from an Inquiry Game that was structured and collaborative rather than spontaneous and individualized. Students and chaperones in all conditions reported enjoying the experience. These results mirror those found in a previous study in which family groups learned the Inquiry Games.

Do students learn science when they go on a field trip to a science museum? Museum educators and researchers have been wrestling for decades over the
question of whether field trips offer significant learning experiences for children, particularly compared to the learning experiences of school (for reviews of the literature, see Bitgood, 1989; DeWitt & Storksdieck, 2008; Koran, Koran, & Ellis, 1989; Price & Hein, 1991).

The question arises in part because the sociocultural contexts of learning in schools and museums differ to a large degree (Falk & Dierking, 1992, 2000; Griffin & Symington, 1997). Schools, driven by the pressures of high-stakes testing and other measures of student achievement, rely largely on students’ extrinsic motivation to receive adult approbation and high grades. Learning activities are often constrained by needs for curriculum coverage of science content, and teachers meet these needs by designing their classrooms to be structured environments, often emphasizing symbolic representations over experimentation with real phenomena (National Research Council, 2007). In a complementary way, science museums are designed to support voluntary, self-directed learning, emphasizing affective responses such as positive attitudes toward science, interest in scientific careers, and feelings of empowerment to make sense of the natural world (Association of Science-Technology Centers, 2002; Ecsite-uk, 2008; Falk & Dierking, 1992, 2000; Friedman, 2008; Hein, 1998; National Research Council, 2009). “Rich with real-world phenomena, these are places where people can pursue and develop science interests, engage in science inquiry, and reflect on their experiences through sense-making conversations” (National Research Council, 2009, p. 15).

The gulf between these two learning contexts, in combination with educators’ epistemological beliefs about what constitutes learning, leads many teachers and museum staff to overlay on open-ended field trip experiences additional structures such as worksheets or didactic tours that emphasize concept learning. Unfortunately, activities that focus on factual learning may actually interfere with student-driven experimentation in museums (Cox-Petersen, Marsh, Kisiel, & Melbe, 2003; Griffin & Symington, 1997; Price & Hein, 1991). At the same time, too little structure can also be detrimental to the field trip as a learning experience by encouraging little or no connection back to the activities of the classroom (Griffin & Symington, 1997; Tal, Bamberger, & Morag, 2005). Research suggests that the most effective field trips are those with intermediate levels of structuring, offering students limited choices (Bamberger & Tal, 2007) or a mix of free-choice and more structured activities (National Research Council, 2009; Price & Hein, 1991).

Some experimental approaches have been successful at fostering concept learning by creating such experiences. For instance, Griffin (1998a, 1998b) developed the SMILES program, in which teachers learn to use the museum as a science resource for their classroom, situating the field trip within a related curricular unit but letting students explore the museum in search of information and ideas relevant to their own projects. Others have developed worksheets and chaperone
guides that support free-choice exploration of school topics (Burtnyk & Combs, 2005; Mortensen & Smart, 2007).

We hypothesize that one of the obstacles to the widespread use of these approaches is that they require some level of preparation—substantial in the case of SMILES, but still significant for teachers using free-choice worksheets and guides. Unfortunately, studies have found that preparatory hurdles are often too great for harried teachers to overcome (Burtnyk, 2004).

Here we report an approach to deepening students’ learning and creating bridges between school-based and informal learning but without such a significant time investment by teachers. Our framework shifts the learning goals for the field trip away from promoting conceptual change and toward helping students build scientific thinking skills. Facilitating skill building presents several advantages as the focus of educational field trips:

1. Critical thinking and scientific inquiry skills are highly valued in documents that define science education at the national level. For example, the National Science Education Standards declare, “Inquiry is central to science learning. When engaging in inquiry, students describe objects and events, ask questions, construct explanations, test those explanations against current scientific knowledge, and communicate their ideas to others” (National Research Council, 1996, p. 2). Similarly, inquiry skills are advocated in both formal and informal settings as a key “strand” of science learning (National Research Council, 2007, 2009). Our goal is to develop field trip programs that engage students in this key aspect of scientific proficiency.

2. Museums provide ideal environments for learning and practicing inquiry skills. While playing with exhibits, students on field trips can try various experiments, make observations, and have memorable experiences (Gottfried, 1980). Over the past 15 years, science museums have further increased their capacity to support such experiences by creating more exhibits that support and extend visitors’ inquiry (e.g., Bailey, Bronnenkant, Kelley, & Hein, 1998; Borun et al., 1998; Humphrey & Gutwill, 2005; Sauber, 1994). By providing an environment that explicitly supports hands-on, direct investigation of natural phenomena, museums offer teachers a vast inquiry learning resource difficult to reproduce in schools.

3. There is some evidence that inquiry skills may transfer more easily than scientific concepts and principles. For example, learning skills such as asking questions can set students up to learn more effectively in new situations (Bransford & Schwartz, 1999).

4. A focus on inquiry skills alleviates the need for teachers to identify specific museum content that intersects with their core school curricula, reducing the call for extensive preparation.
In summary, we argue that skill building may bridge the formal and informal worlds of science education by using the strengths of science museums to meet a significant need in schools. We are not the first or only researchers to promote the idea of inquiry learning in field trips. DeWitt and Storksdieck’s (2008) review of field trip effectiveness concluded that “field trip offerings should be based on exploration, discovery and process skills rather than transmission of facts” (pp. 190–191). In a review of their own evaluations of field trip programs spanning 30 years, Price and Hein (1991) defined *educationally effective programs* as “those in which products are not emphasized, inquiry is sparked, open-ended questions are generated, and students actively participate and appear involved” (p. 510).

What our current work contributes is a set of specific programmatic techniques for giving field trip students a “crash course” in inquiry skills within the context of a museum’s exhibit floor. Learning or even practicing inquiry skills on a field trip is not a simple endeavor. Although open-ended exhibits can support inquiry, visitors, especially children, often do not conduct coherent, in-depth investigations to answer their questions (Randol, 2005). Some inquiry skills are challenging even for people with science backgrounds (Allen, 1997; Loomis, 1996).

In a previous study, we explored related issues in the context of the general public, creating a set of Inquiry Games for family groups that were well received and that successfully enhanced certain inquiry practices, compared to controls, when families played them at a novel exhibit (Allen & Gutwill, 2009; Gutwill & Allen, 2010a). Based on this success, we have attempted to adapt the Inquiry Games for use by field trip groups, attending to key differences between family and field trip learners, to determine whether they could serve as useful models for furthering the educational agenda of museum field trip visitors.

Funded by the National Science Foundation, this project was carried out at the Exploratorium, San Francisco’s museum of science, art, and human perception, and was named GIVE (Group Inquiry by Visitors at Exhibits).\(^1\)

**RESEARCH QUESTIONS**

Our study of Inquiry Games with field trip groups addressed the following questions:

- Can inquiry be taught in the museum context without the need for pre- or postvisit work in the classroom?

\(^1\)The GIVE team developed the Inquiry Games in collaboration with staff from the Exploratorium’s Explainer and Institute for Inquiry programs, the Visual Thinking Strategies program staff at the San Francisco Museum of Modern Art, and an advisory board of experts with a range of expertise.
• Can such coaching of inquiry skills meet participants’ expectations for the field trip and provide an optimal mix of structure versus choice, as suggested in the literature?
• Is there any evidence that the inquiry skills learned during such a field trip experience transfer to new settings?

In addition, we addressed a comparative question related to learners:

• How does the inquiry learning of field trip groups compare with that of intergenerational family groups?

DESIGNING THE INQUIRY GAMES

Skills
Most research-based inquiry programs in schools aim to teach skills such as asking questions, making predictions, designing experiments, analyzing data, reasoning with models, drawing conclusions, and communicating results (e.g., Minstrell & van Zee, 2000; White & Frederiksen, 1998). During our pilot phase, we attempted to help family groups learn such skills in the form of a sequence and then assessed whether they could apply them at a novel exhibit. Unfortunately, in the museum context the cognitive load proved too great for the families; groups would simply forget and thus skip certain skills. (The details of our pilot work in this area are reported in Allen & Gutwill, 2009.) Based on the results of several iterations, we limited the inquiry skills in our program to the following two:

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

These two skills complement students’ natural exploration activity at exhibits, are intellectually accessible to diverse groups of students aged 10–13, and are simple enough for students to understand quickly and remember easily. Previous research has shown that these skills are rarely practiced by intergenerational (family) visitors (Randol, 2005). We hypothesized that school students would be even less adept at applying them, especially in field trip groups in which the child-to-adult ratios are higher than in families. Furthermore, even if students do silently ask questions before manipulating an exhibit or mentally draw conclusions after using it, articulating such thoughts overtly can move a group’s investigation forward by putting ideas into public spaces for improvement by peers (Quintana
et al., 2004; Scardamalia, 2002). We further hoped that these skills might foster other inquiry behaviors, such as explanation building or hypothesis testing.

Many inquiry processes developed for schools take the form of a cycle that students may repeat as results lead to new questions (e.g., Champagne, Kouba, & Hurley, 2000; Chinn & Malhotra, 2002; Songer, 2004; White & Frederiksen, 1998). Similarly, our inquiry process, shown in Figure 1, allows for cyclical investigation of phenomena at museum exhibits. The dark ovals in Figure 1 indicate the skills that we explicitly taught visitors, whereas the light oval represents a skill that visitors spontaneously perform when using an interactive science museum exhibit (Randol, 2005).

Pedagogical Structures

In our prior work, the research team had created activity structures referred to as Inquiry Games in order to teach visiting families the two targeted inquiry skills. The Inquiry Games did not have winners and losers but were “games” in the sense that they contained steps for play, rules, and a structure for social interaction. The term game helped us focus the attention of students and chaperones on the inquiry process rather than on the exhibits used in the study.

FIGURE 1  Simplified inquiry cycle with the two targeted skills shown in the dark ovals.

---

2The Inquiry Games did not have winners and losers but were “games” in the sense that they contained steps for play, rules, and a structure for social interaction. The term game helped us focus the attention of students and chaperones on the inquiry process rather than on the exhibits used in the study.
When switching from families to field trip groups, we felt that the formats of the two Inquiry Games required only small modifications to be useful.

**Juicy Question game.** In the first game, called Juicy Question, students and their chaperone work 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 initially knows the answer. To begin, the group members explore the exhibit to become familiar with it. Next, they stop and take turns sharing a question they each have about the exhibit. They then choose one of the questions and investigate it as a group. Finally, they stop to reflect on what they have discovered, sharing their ideas until they feel finished. Each person receives a card to remind them of the two skills: asking juicy questions (PA) and sharing discoveries (IR).

The Juicy Question game is highly collaborative: Students and their chaperone must negotiate to choose a single question to pursue, and they must agree on when to stop experimenting to generate interpretations or new questions. Managing the collaborative process requires a facilitator. The museum educator takes this role at first but then asks the chaperone to take it over, using a process of scaffolding and fading (Brown, Collins, & Duguid, 1989; Collins, Brown, & Newman, 1989). Because chaperones have varying degrees of facilitation experience, they are given a special card with steps for running the game. In the current study, each field trip group played the game twice (at two different exhibits) before they were asked to do it alone at the posttest exhibit (see Figure 2 for the experimental design).

**Hands Off game.** We hypothesized that the Juicy Question game might require more collaboration than a field trip group would be capable of achieving, given that chaperones are often unfamiliar with most of the students in their care and that we could not count on an established dynamic of collaboration among the students themselves. We anticipated that students might respond more naturally to an activity structure that supported more individual activity. Our alternative game, Hands Off, allows anyone at any time to call out “Hands off!”, at which time the other group members must stop using the exhibit and listen to the caller. At that point, 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. Overall, the game teaches students and chaperones to use the same two inquiry skills as Juicy Question but in a more individual and spontaneous way. No facilitator is needed, and the game is particularly easy to remember. For consistency with Juicy Question in our study, the chaperone was given a special card as a reminder to support the rules of the game.
Principles Underlying the Inquiry Games

Despite their different pedagogical structures, both Inquiry Games incorporated several key principles from the school-based science education literature.

Build on learners' prior knowledge. It is well known that students build new knowledge out of prior experience (Piaget, 1978; Roschelle, 1995; von Glasersfeld, 1989; Vygotsky, 1962). Both Inquiry Games incorporated this principle by allowing students to raise and pursue their own questions about an exhibit rather than prescribing specific science content that should be learned. This also helped support students of different ages, from different schools, studying different science topics, to work in their own “zones of proximal development” (Vygotsky, 1978). According to DeWitt and Storksdieck (2008), field trip programs struggle to reach field trip students at their individual levels: “It can be difficult for museum practitioners to provide experiences specifically appropriate to each student’s prior knowledge” (p. 185). By generating and pursuing their own questions, students could better connect with the exhibit at their own level.

Teach via modeling, scaffolding, and fading. We embraced the 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 et al., 1989; Vygotsky, 1978). In the current study, students and chaperones played an Inquiry Game three times, each time with a novel exhibit. The educator gradually faded from the role of facilitator, first by helping the chaperone facilitate, then by disappearing entirely and giving the group full autonomy. We envisioned this process as embodying nested zones of proximal development in which first our educator scaffolded the process for chaperones and students and later chaperones scaffolded it for students. Such an approach was particularly important for helping parent chaperones who were assigned to a group of children they may not have known well (Parsons & Breise, 2000).

Identify skills explicitly. Students learn new skills better when they are explicitly articulated, demonstrated, and practiced (Labudde, Reif, & Quinn, 1988; Palincsar & Brown, 1984; White & Frederiksen, 1998). In our study, the educator explicitly stated the purpose and key steps of the Inquiry Game and named instances of the specific skills as students practiced them in the context of the conversation at the exhibit (e.g., “That sounds like a discovery,” or “Does anyone else have a question they want to ask?”).
Support metacognition. Another important factor in learning is metacognition, or learners’ abilities to monitor and reflect on their own understanding (Bransford, Brown, & Cocking, 2003; Brown, 1975; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Flavell, 1973). In this project, we targeted two skills that had a strong metacognitive component: asking a question and reflecting on what was learned during an investigation. Both skills required students and chaperones to monitor their own knowledge state and its changes.

Support collaboration. Finally, learning in groups often improves motivation and achievement (Tobin, Tippins, & Gallard, 1994). This is especially important in museums, where field trip students visit in class groups (Griffin, 1998b). Our study focused on field trip groups of 5–7 students and a parent chaperone who had probably never before worked together. Both Inquiry Games offered turn-taking strategies to help the participants learn effectively as a group.

In addition to these learning principles, we incorporated three principles from the literature on museum field trip groups.

Strike a balance between choice and guidance. In their review of research on museum field trips, DeWitt and Storksdieck (2008) concluded that “field trips should provide a moderate amount of structure while still allowing for free exploration” (p. 186). Our Inquiry Games attempted to find that sweet spot: On the one hand, they structured students’ interactions by encouraging them to ask questions, make interpretations, and take turns in rule-based ways. On the other hand, the games allowed students the freedom to choose which exhibits to use and to generate and pursue their own questions at the exhibits. In short, we designed the games to help students improve their own collaborative inquiry process within a free-choice learning environment.

Place realistic demands on teachers. Much of the literature on museum field trips laments teachers’ lack of integration of the visit into the classroom learning experience (e.g., Cox-Petersen et al., 2003; Griffin & Symington, 1997). Preparing students and chaperones for an educational museum experience can be difficult, as many teachers do not visit the museum before the field trip (Tal et al., 2005) and others may not understand how learning works in museums (Kisiel, 2003). After the field trip, many teachers fail to connect the experience back to students’ work in the classroom (DeWitt & Storksdieck, 2008; Griffin & Symington, 1997). The Inquiry Games, learnable in less than 20 min at two museum exhibits, required no preparation from students, chaperones, or teachers.3

3Participants were required to do additional administrative preparation, as described later, because they were participating in a research study, but the games themselves could be learned without the need for prior preparation.
Create a useful role for chaperones. Few studies have focused on the role of chaperones, but those that have have found that chaperones can have an important impact on student learning (Burtnyk & Combs, 2005; Parsons & Breise, 2000). Unfortunately, the potential for chaperones to make an educational contribution is often squandered by the overwhelming need for logistical support during the field trip (Parsons & Breise, 2000). In our program, chaperones learned, practiced, and fulfilled the role of group facilitator in the course of playing the games. The facilitator role offered chaperones the opportunity to participate with students in a useful learning process.4

STUDY DESIGN

After studying the family group audience, we conducted a similar randomized controlled study to test the impact of the Inquiry Games on field trip groups’ inquiry behaviors at exhibits. (For more details about the experimental design, see Gutwill & Allen, 2010a.) We randomly assigned participating field trip groups 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 activity the field trip group engaged in:

1. Juicy Question: The group learned and played two rounds of the Juicy Question game at two different exhibits.
2. Hands Off: The group learned and played two rounds of the Hands Off game.
3. Exhibit Tour Control: The group listened to an interactive description of the science content and developmental history of two exhibits but was not taught any generalizable skills.
4. Pure Control: The group used the exhibits without any game or educational mediation.

We studied 46 field trip groups in each of the four conditions, or 184 groups altogether. We recruited field trip groups by contacting teachers from public schools who had already reserved a field trip to the museum.5 Amenable teachers

4We acknowledge that this program took place within the specialized context of a research study in a lab separated from the museum floor and that all parties had agreed to participate ahead of time. Our future work will explore the degree to which chaperones can facilitate the games on the open floor.
5We recruited groups from public schools and used published data on their free and reduced lunch programs as a rough measure of students’ socioeconomic status. In the 113 schools that sent student groups to participate in our study, an average of 49% of the students qualified to receive free or reduced lunch. We did not, however, collect socioeconomic data on individual students.
arrived at the museum with parent-signed consent forms for each child and pre-selected participant groups composed of 5–7 students and one parent chaperone. (Teachers were not allowed to act as chaperones out of concern that they would enforce their own learning agendas rather than implement those in the study. We regarded this as a reasonable filter on our sample, given that most of the children on field trips are entrusted to chaperones who are not themselves teachers.) The study was conducted in a research laboratory at the back of the museum so that we could videotape the groups using exhibits for detailed analysis (see Figure 3).
By comparing students and chaperones who had learned the Inquiry Games to those who had not, we could determine whether the games helped field trip groups conduct in-depth inquiries at novel exhibits.

Data Collection

The sequence of participation for each field trip group was as follows: Students and their chaperone entered the lab, 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, the chaperone and one child were chosen to participate in interviews after the experience (see Figure 2).\(^6\)

\(^6\)To reduce distraction and group fragmentation in the lab, we covered the exhibits with tablecloths and removed them only for the exhibit in use at any given time. In addition, group members were asked to remain together at each exhibit until the entire group was ready to move on to the next exhibit.
Exhibits in the Study

The four exhibits used in the study were originally developed to support visitor-driven investigation by (a) offering multiple options for visitors to explore, (b) having no obvious endpoint or message to convey, and (c) providing multiple access points so that group members could use the exhibit simultaneously (Humphrey & Gutwill, 2005). The exhibits, in the order field trip groups used them, were as follows:

1. *Shaking Shapes (pretest exhibit)*. This exhibit consists of a vibrating table upon which loose geometrical shapes spin and shake. Student groups typically build structures and investigate their structural stability and rotational motion. Groups in our study used this exhibit without any prior instruction (see Figure 4).

2. *Floating Objects (second exhibit)*. Students can experiment with air pressure and flow by placing differently shaped objects—wiffle balls, small basketballs, plastic pears and apples, and ping-pong balls—in a vertical air stream. By tilting the air jets, groups may discover the surprising result

![FIGURE 4](image-url) A field trip group using the Exploratorium’s Shaking Shapes exhibit.
that a ball can float in a slanted air stream. Groups in the inquiry conditions were facilitated by our staff educator. Figure 5 shows the Floating Objects exhibit.

3. Unstable Table (third exhibit). By building structures on a gimbaled table, students explore the concepts of torque and counterweight to keep the platform balanced. Chaperones in the two Inquiry Game conditions were asked to try facilitating while our staff educator provided support (see Figure 6).

4. Making Waves (posttest exhibit). Magnetically coupled pendulums hang from a common spine, creating rippling wave patterns or more chaotic motion, depending on students’ actions. Students may experiment with wave phenomena and magnetic attraction. Groups used this exhibit without the presence of our staff educator but were asked to play an Inquiry Game if they had learned one. See Figure 7 for the Making Waves exhibit.

All field trip groups encountered the exhibits in the fixed sequence presented above. We had considered using a counterbalanced design but rejected it because our resources did not permit the increase in sample size needed to achieve the same degree of analytical power. In addition, our experiment already utilized
two blocking variables (teacher and condition) and blind recruiters, and we were concerned that adding another blocking variable (exhibit sequence) would introduce substantial human error in managing the study. The fixed design maximized power for identifying the differential treatment effects of our program, with the limitation that those effects may not generalize beyond this particular exhibit sequence.

Assessing Inquiry

Using Studiocode™ video analysis software, we assessed groups’ inquiry behaviors at the pretest and posttest exhibits to determine whether playing the Inquiry Games enhanced their inquiry skills. We used the same dependent variable codes to assess field trip groups’ inquiry behaviors as we had used with our family groups, because we felt the behaviors would be common to both studies. These were as follows.

Engagement. We measured the length of time a group chose to spend at the exhibit (i.e., holding time) as an indicator of its engagement.
Proposing Actions (Skill 1). We captured the number, frequency, and duration of instances by groups of PA, which was one of the skills targeted in the intervention. The PA utterances were coded at two levels—high and low. A low-level PA (PA1) would include only the action a participant wanted to take (“What if we pull it faster?”) or only the effect they hoped to induce (“Can we make the wave bigger?”), but not both. A high-level PA (PA2) would include both the action and the desired result (“Let’s see if we get the end magnets to swing hard by only pushing the middle magnets”).

Interpreting Results (Skill 2). As with PAs, we counted the number, frequency, and duration of instances of IR. We coded IR at two levels. A low-level IR (IR1) would involve only direct observation, with very little abstraction beyond
the moment (“The yellow ones barely spin”). A high-level IR (IR2) would generalize (“If you stack them up, they spin faster”), explain the results of the experiment (“Those shapes spin faster because they barely touch the table so there’s less friction”), or offer an analogy (“It’s like the way dominoes fall down.”)

**Collaborative Explanations.** We tallied the number of times multiple group members interpreted results in rapid succession and called these “Consecutive IRs.” An example would be “The papers we put between the pendulums didn’t stop them” followed immediately by “Yeah, because magnets can go through paper.” Consecutive IRs often occurred in strings of various length. For this code, we counted the percentage of a group’s IRs that were consecutive, as well as the longest string of Consecutive IRs spoken in a group. Typically such conversational turn taking indicated that group members were building a shared understanding. We did not explicitly teach groups to discuss their interpretations in this way but assessed it on the grounds that developing an explanation is an important aspect of successful inquiry, and building explanations jointly with others reflects a collaborative process.

**Coherent Investigations.** To assess the “depth” of a group’s inquiry, we counted the number of times group members connected their investigations so that one experiment followed another in a coherent way (e.g., “Let’s see what happens when you block out a magnet” followed by a second experiment proposed with “If we block out a lot of magnets, what happens?”). This meant looking at each PA instance to see if it was linked to the one before it by a common theme (e.g., interrupting magnetic field). We counted the number of PAs that were linked in this way and called the longest string of them the Linked PA score. Like Consecutive IRs, Linked PAs was not a skill targeted in the games but was taken as an additional and independent indicator of the overall quality of the group’s inquiry.

**Reliability of Inquiry Coding Schemes**

Multiple research assistants blind either to condition or to the purpose and hypotheses of the study coded the videos of field trip groups at the pretest and posttest exhibits. In our previous study of families, we measured the interrater reliability of our coding schemes and found that they surpassed the acceptable norms reported in the literature. However, we were concerned that field trip groups might be more difficult to code than families. (For example, field trip groups are larger

---

7For 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 s. Because of the rules of our coding scheme, this ensured that two different people uttered the two IRs.
and have more children, which could make consecutive comments harder to discern.) Consequently, we retested the reliability of the coding schemes on 10% of the new data set using measures appropriate to the nature of the specific codes. For the PA and IR codes, we calculated 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 0.77, about the same as the kappa for the codes of the family group data (0.76). (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 an intraclass correlation coefficient assessing the degree to which the coding scheme distinguished among field trip groups equally well for different coders (Bakeman & Gottman, 1997). The coefficient for Linked PA codes was found to be 0.68, which is conventionally viewed as “good” reliability (Cicchetti & Sparrow, 1981; Fleiss, 1981). (This was slightly lower than the statistic found for the codes of the family data, 0.70.) We calculated scores for Consecutive IRs (Collaborative Explanations) automatically using the IR data, so they did not necessitate an additional check for interrater reliability. Likewise, interrater reliability was not necessary for time at exhibit (engagement), because frame-by-frame measurements were easily accurate to within 1 s.

Participants’ Self-Reports

To assess participants’ responses to the experience, we interviewed and surveyed the chaperone and one randomly chosen child from each group. We conducted two types of interviews. Immediately after the participants used the exhibits, we asked them what they liked and disliked about the experience and how likely they would be to use the Inquiry Game (or the Exhibit Tour information) in the future. Three weeks later, we interviewed the same child and chaperone again by phone, asking what they could remember of the experience and whether they had in fact used the game (or Exhibit Tour information) on the museum floor or outside the museum. All interview data were coded redundantly by two coders who negotiated any disagreements.

A demographic survey administered in the lab asked all students and chaperones about their age, gender, highest level of schooling, and highest level of science attained in school. The survey data were used in the analysis to ensure that there were no demographic differences across treatment and control conditions.

Assessing Conceptual Understanding

Although we developed the entire program to deepen scientific inquiry rather than teach any specific content, we were interested to see whether students’ 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” students and chaperones in the interview, so
we indirectly measured correctness of understanding by revisiting the video data
and recoding all sophisticated interpretations (IR2s) in terms of their fit with sci-
entific canon. Two coders separately categorized each IR2 as correct, incorrect, or
uncodable and then met to work out all disagreements. An uncodeable response
would be one for 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.

Hypotheses and Planned Comparisons

Before collecting our data, we articulated our hypotheses for the study. We
believed that the Inquiry Games would have similar effects on family and field
trip groups, so we decided to conduct the same planned comparisons on field
trip group data that we had performed on family group data in our previous study.
(Carrying out identical analyses in the two studies also had the advantage of allow-
ing us to compare results across the two audiences.) We had three overarching
hypotheses for the field trip groups, with concomitant planned comparisons:

1. Inquiry Games should improve inquiry behaviors. As with families, we
predicted that students and chaperones who learned the Inquiry Games
would use the inquiry skills we taught them (PA and IR) more fre-
cently and for longer durations. We thought it possible but less likely
that they would also outperform the control groups in their frequencies
of other inquiry behaviors that were not directly taught (e.g., Coherent
Investigations and Collaborative Explanations). Overall, our planned com-
parison predicted an effect of inquiry, such that the Juicy Question and
Hands Off conditions would outperform the Exhibit Tour condition on our
measures of inquiry.

2. Prior mediation should enhance the experience. Based on studies showing
that interactions with museum staff are among the most memorable aspects
of a museum experience (e.g., Allen, 2004; Piscitelli & Weier, 2002),
we predicted that students and chaperones who had spent time with our
museum educator would have a more positive experience than those who
had not. We also thought that prior mediation might lead to better inquiry
behaviors at the posttest exhibit, particularly longer holding times. Thus,
our planned comparison predicted an effect of prior mediation, such that
the Exhibit Tour condition would outperform the Pure Control condition.

3. Pedagogy should 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 might make for deeper inquiry, whereas the focus of Hands Off on turn taking might make it easier for students to work together. Consequently, our planned comparisons used two-tailed measures to test for an effect of pedagogy.

To implement the 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 using analyses of variance. When differences were significant for the planned comparisons, we conducted additional post hoc t tests to reveal underlying patterns.

RESULTS

In this section, we first report on the demographics of the students and chaperones in the different conditions. Then we describe the effects of the Inquiry Games on groups’ inquiry behaviors. Next we report participants’ reflections on their experiences. Finally, we describe the scientific correctness of the groups’ interpretations.

Overall, the main pattern of results was as follows: (a) There were no differences in demographics across conditions; (b) the Inquiry Games, especially Juicy Question, increased the quantity and quality of field trip groups’ scientific inquiry behaviors from pretest to posttest; (c) although all students enjoyed the experience, a quarter of the students who had learned Juicy Question, more than the number of students in any of the other conditions, reported using the targeted skills at museum exhibits outside the lab; and (d) most groups in the two inquiry conditions tended to make correct interpretations after instruction.

Demographic Attributes of the Field Trip Groups

Students and chaperones in each group filled out a demographic survey before interacting with the exhibits; the results are shown in Tables 1 and 2. Chi-square and Fisher exact tests of the data indicated that there were no differences by condition for any of the variables we assessed. This result means that any differences in inquiry behaviors should not be attributed to differences in the measured demographic variables.

Although we found no differences in demographics across conditions, we hoped to use the demographic variables of age and education level as covariates in our planned comparisons. However, only a small number of dependent variables correlated with these demographic variables. Even in those cases, using
### TABLE 1
Demographics of Student Groups in Each Experimental Condition

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Juicy Question</th>
<th>Hands Off</th>
<th>Exhibit Tour</th>
<th>Pure Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eldest student’s age</td>
<td>11.1 (0.72)</td>
<td>11.3 (0.78)</td>
<td>11.1 (0.64)</td>
<td>11.1 (0.73)</td>
</tr>
<tr>
<td>Group grade level</td>
<td>5.2 (0.59)</td>
<td>5.4 (0.62)</td>
<td>5.3 (0.59)</td>
<td>5.2 (0.58)</td>
</tr>
<tr>
<td>Number of boys in group</td>
<td>2.3 (1.6)</td>
<td>2.7 (1.7)</td>
<td>2.6 (1.5)</td>
<td>2.8 (1.7)</td>
</tr>
<tr>
<td>Number of girls in group</td>
<td>3.2 (1.6)</td>
<td>2.9 (1.5)</td>
<td>2.9 (1.5)</td>
<td>3.0 (1.3)</td>
</tr>
<tr>
<td>Proportion of students in group with “special interest, knowledge, or training” in science</td>
<td>0.46 (0.32)</td>
<td>0.51 (0.30)</td>
<td>0.55 (0.27)</td>
<td>0.52 (0.31)</td>
</tr>
</tbody>
</table>

### TABLE 2
Demographics of Parent Chaperones in Each Experimental Condition

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Juicy Question</th>
<th>Hands Off</th>
<th>Exhibit Tour</th>
<th>Pure Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–25</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>26–35</td>
<td>9</td>
<td>8</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>36–45</td>
<td>27</td>
<td>18</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>46–55</td>
<td>8</td>
<td>13</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>56–65</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>13</td>
<td>15</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Female</td>
<td>33</td>
<td>31</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>Highest schooling level attained</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No response</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>High school</td>
<td>10</td>
<td>12</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>College/vocational school</td>
<td>21</td>
<td>17</td>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td>Graduate school</td>
<td>9</td>
<td>11</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Highest STEM level attained</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Middle/high school</td>
<td>16</td>
<td>21</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>College classes</td>
<td>23</td>
<td>19</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>College degree</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Chaperones with “special interest, knowledge, or training” in science</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>18</td>
</tr>
</tbody>
</table>

*Note.* Data are frequencies. STEM = Science, Technology, Engineering and Math education.
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

Our analyses indicated that although Hands Off was helpful, the Juicy Question game was best at improving the inquiry skills of students and chaperones at a novel exhibit. Table 3 shows the mean pre/post difference scores for the inquiry behaviors we coded in the video data, and Table 4 shows the results of the planned and post hoc comparisons conducted on these scores. Planned comparisons revealed a clear effect of inquiry: The two inquiry conditions repeatedly outperformed the mediated control condition. The effect sizes for inquiry, represented by Cohen’s $d$ statistic, ranged from 0.3 to 0.8. The effect of pedagogy was a little less consistent: The Juicy Question condition outperformed Hands Off in five respects, but Hands Off showed superior performance in one. The size of the pedagogy effect ranged from 0.4 to 0.7. We found a significant effect of prior mediation ($d = 0.4$) on only 1 of 13 measures, suggesting that the Exhibit Tour had little effect on groups’ inquiry behavior at a novel exhibit.

Both Games Improved Inquiry, but Juicy Question Performed Better

Both Inquiry Games improved students’ inquiry behaviors as compared to the Exhibit Tour. Still, there were differences in the ways in which Juicy Question and Hands Off affected field trip groups. We next describe the impact of the Inquiry Games on each of the dependent variables measured.

Engagement. There were no group differences in the pre/post change in holding time at the exhibits (see Figure 8). We expected to see an increase in holding time for groups in the two inquiry conditions, especially because we had found an effect of inquiry in the previous study of family groups. Instead, field trip groups apparently became better at inquiry (see the variables below) but did not spend significantly more time doing it.

Skill 1: Proposing Actions. Groups in both inquiry conditions significantly increased the duration and sophistication of their PA utterances compared with groups in the Exhibit Tour condition (see Tables 3 and 4, where “sophistication” is measured as PA2s as a percentage of PAs). Groups in the Juicy Question condition also significantly increased the number of PAs they made compared to Hands Off.

---

8 All graphs show pretest and posttest scores to emphasize the pre/post design of the study. However, our analyses compared groups based on their pre/post difference scores, thus accounting for any apparent pretest disparities across condition.
### TABLE 3
Pre/Post Difference Scores for Inquiry Behaviors Captured on Video

<table>
<thead>
<tr>
<th>Inquiry Behavior</th>
<th>Measure</th>
<th>Juicy Question</th>
<th>Hands Off</th>
<th>Exhibit Tour</th>
<th>Pure Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement</td>
<td>Time at exhibit (minutes)</td>
<td>1.8 (3.9)</td>
<td>0.41 (5.3)</td>
<td>0.09 (2.5)</td>
<td>0.17 (3.6)</td>
</tr>
<tr>
<td>Skill 1: PA</td>
<td>Number of PAs</td>
<td>6.7 (23)</td>
<td>−6.1 (22)</td>
<td>−1.02 (14)</td>
<td>1.02 (13)</td>
</tr>
<tr>
<td></td>
<td>PAs per minute</td>
<td>0.18 (1.8)</td>
<td>−0.56 (1.7)</td>
<td>−0.09 (1.7)</td>
<td>0.11 (1.4)</td>
</tr>
<tr>
<td></td>
<td>Duration of PAs (minutes)</td>
<td>0.93 (1.2)</td>
<td>1.4 (1.2)</td>
<td>0.34 (0.71)</td>
<td>0.43 (0.73)</td>
</tr>
<tr>
<td></td>
<td>PA2s as a percentage of PAs</td>
<td>6% (10%)</td>
<td>6% (15%)</td>
<td>1% (9%)</td>
<td>3% (8%)</td>
</tr>
<tr>
<td>Skill 2: IR</td>
<td>Number of IRs</td>
<td>15 (23)</td>
<td>2.6 (27)</td>
<td>4.0 (16)</td>
<td>−0.80 (25)</td>
</tr>
<tr>
<td></td>
<td>IRs per minute</td>
<td>0.96 (2.2)</td>
<td>0.4 (1.7)</td>
<td>0.59 (2.0)</td>
<td>−0.23 (2.5)</td>
</tr>
<tr>
<td></td>
<td>Duration of IRs (minutes)</td>
<td>0.82 (1.02)</td>
<td>1.2 (1.4)</td>
<td>0.32 (0.95)</td>
<td>0.36 (1.07)</td>
</tr>
<tr>
<td></td>
<td>IR2s as a percentage of IRs</td>
<td>13% (10%)</td>
<td>12% (16%)</td>
<td>4% (10%)</td>
<td>5% (10%)</td>
</tr>
<tr>
<td>Collaborative Explanations</td>
<td>Consecutive IRs as a percentage of IRs</td>
<td>3% (13%)</td>
<td>5% (12%)</td>
<td>0% (12%)</td>
<td>0% (11%)</td>
</tr>
<tr>
<td></td>
<td>Consecutive IR string length</td>
<td>0.74 (2.08)</td>
<td>0.67 (1.7)</td>
<td>0.17 (1.1)</td>
<td>−0.11 (1.6)</td>
</tr>
<tr>
<td>Coherent Investigations</td>
<td>Linked PA string length</td>
<td>2.1 (4.06)</td>
<td>0.78 (3.7)</td>
<td>0.52 (3.02)</td>
<td>0.00 (3.1)</td>
</tr>
<tr>
<td>Other</td>
<td>Reads exhibit label</td>
<td>1.6 (1.7)</td>
<td>0.76 (1.5)</td>
<td>1.6 (1.9)</td>
<td>1.4 (1.9)</td>
</tr>
</tbody>
</table>

**Note.** Juicy Question = structured Inquiry Game; Hands Off = spontaneous Inquiry Game; Exhibit Tour = mediated control; Pure Control = unmediated control; PA = Proposing Actions; PA2 = high-level Proposing Actions; IR = Interpreting Results; IR2 = high-level Interpreting Results.
TABLE 4
Planned and Post Hoc Comparisons for Inquiry Behaviors Captured on Video

<table>
<thead>
<tr>
<th>Inquiry Behavior</th>
<th>Measure</th>
<th>Planned Comparisons (ANOVA), $F(1, 180)$</th>
<th>Post Hoc t Tests, $t(90)$, $p &lt; .05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(Cohen's $d$)</td>
<td></td>
</tr>
<tr>
<td>Engagement</td>
<td>Time at exhibit (minutes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill 1: PA</td>
<td>Number of PAs</td>
<td>11 (0.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PAs per minute</td>
<td>4.7 (0.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duration of PAs (minutes)</td>
<td>21 (0.8)</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>PA2s as a percentage of PAs</td>
<td>7.3 (0.5)</td>
<td>2.7</td>
</tr>
<tr>
<td>Skill 2: IR</td>
<td>Number of IRs</td>
<td>7.7 (0.6)</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>IRs per minute</td>
<td>3.6 (0.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duration of IRs (minutes)</td>
<td>13 (0.7)</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>IR2s as a percentage of IRs</td>
<td>16 (0.7)</td>
<td>4.3</td>
</tr>
<tr>
<td>Collaborative</td>
<td>Consecutive IRs as a percentage</td>
<td>2.9 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Explanations</td>
<td>of IRs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consecutive IR string length</td>
<td>3.1 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Coherent</td>
<td>Linked PA string length</td>
<td>2.2 (0.3)$^a$</td>
<td>3.4 (0.4)$^a$</td>
</tr>
<tr>
<td>Investigations</td>
<td>Other</td>
<td>4.9 (0.5)</td>
<td>-2.4</td>
</tr>
</tbody>
</table>

Note. ANOVAs for the effects of prior mediation and of inquiry used one-tailed tests; all other comparisons used two-tailed tests. Unless otherwise noted, only effects significant at the .05 level are shown. Blank cells indicate no significant differences. ANOVA = analysis of variance; ET = Exhibit Tour; PC = Pure Control; JQ = Juicy Question; HO = Hands Off; PA = Proposing Actions; PA2 = high-level Proposing Actions; IR = Interpreting Results; IR2 = high-level Interpreting Results.

$^a$ Statistically marginal effect ($p < .07$).
groups (see Figure 9 for the change in the number of PAs). However, the Hands Off groups outperformed the Juicy Question groups on the duration of their PA utterances. These results were different in the prior family study, in which the only effect of the inquiry conditions was on duration of PAs, with Juicy Question, not Hands Off, outperforming controls. As an example of a PA becoming more sophisticated, consider that one group’s PA utterances at the pretest exhibit typically focused only on either the intended action or the desired effect but not both, such as, “Let’s put them all in a circle” (intended action). At the posttest exhibit, more of their PAs included both, such as, “I wonder if we just had one going and see if it would like affect the whole thing” (intended action and desired effect).

Skill 2: Interpreting Results. Similar to the pattern with PAs, groups in both inquiry conditions showed a significant increase, compared to the Exhibit Tour groups, in the duration and sophistication of their IR utterances (see Tables 3 and 4, where “sophistication” is measured as IR2s as a percentage of IRs). For example, many interpretations in the pretest were simple observations (IR1s), such as, “It fell,” “It’s like moving around,” and “That’s shaking.” At the posttest exhibit,
more of the interpretations involved explanatory reasoning (IR2s), such as, “The heavier one goes farther” or “It’s the ones that are on the outside move farther back and forth and then the ones in the middle don’t move as far.” Groups in the Juicy Question condition made more IRs than groups in either the Hands Off or Exhibit Tour conditions (see Figure 10 for the change in the number of IRs). These findings are similar to those from the prior family study, in which the Juicy Question condition significantly increased the number, duration, and sophistication of families’ IRs. This variable also revealed the only effect of prior mediation, in which the Exhibit Tour groups increased the number of IRs made per minute more than the Pure Control groups.

**Collaborative Explanations.** Field trip groups in the inquiry conditions outperformed those in the Exhibit Tour Control on the percentage of their interpretations that were consecutive (percent Consecutive IRs) and on the maximum number of Consecutive IRs made in a row (Consecutive IR string length). As an example of consecutive interpretations, consider an exchange in a Juicy Question
group between two girls who were trying to understand how the “force” of a magnet is transferred down the length of the exhibit:

Girl 1: The ones that are over here, they move farther with it—
Girl 2: They move with it and then this one gets smaller and doesn’t move—
Girl 1: Because it’s not as near it, it doesn’t get that much, you know, force.

In this example, the students build on each other’s understanding of the exhibit. Note that Girl 2 first agrees with Girl 1 (“They move with it”) and then extends the explanation (“and then this one gets smaller and doesn’t move”). Girl 1 then continues to build their understanding, starting her next utterance with “Because . . .” This kind of collaborative interpreting is typical of Consecutive IRs.

Despite the significant effect of inquiry on Collaborative Explanations, the post hoc analyses found no differences, showing that the difference surfaced only when we combined the Juicy Question and Hands Off groups together; either condition
on its own did not improve enough over Exhibit Tour to show a significant difference (see Figure 11) The pattern for Consecutive IRs was different for the family groups, with Juicy Question families significantly outperforming Exhibit Tour families on both Consecutive IRs as a percentage of IRs and Consecutive IR string length.

Coherent Investigations. The planned comparisons revealed marginal effects of inquiry and pedagogy for groups conducting long strings of investigations that were linked by a common theme (Linked PAs). The post hoc test\(^9\) clarified that groups in the Juicy Question condition, not those in Hands Off, outperformed the Exhibit Tour groups (see Figure 12). Apparently, the performance

![Figure 11](image)

**FIGURE 11** Mean percentage of Interpreting Results (IR) codes that were Consecutive IRs made by groups in each condition.

\(^9\)Although the planned comparison results were only marginal, we felt that post hoc tests were justified because the family study had showed a significant difference on this variable.
FIGURE 12  Mean number of Linked Proposing Actions (PAs) made by groups across conditions.

of Hands Off groups was washing out the effect from Juicy Question groups in the planned comparisons.

For example, in the Juicy Question condition, one group composed of three girls (all 10 years old), two boys (one 9 and one 10), and a male chaperone became interested in the issue of whether it would be possible to “block” the effect of the magnets in the exhibit. (The exhibit they were using, called “Making Waves,” consisted of 23 pendulums hanging from a horizontal rod. Each pendulum bob had a magnet on one side, so when one pendulum swung, the others started swinging in a wave pattern.) The group tried six different experiments to investigate this single topic.\(^{10}\)

\(^{10}\)Parts of this group’s transcript originally appeared in Gutwill and Allen (2010b).
Boy 1: Let’s see what happens when you block out a magnet.

[Experiment 1: They try holding one pendulum up and out of the way of the others and then move another pendulum to see if it can affect its neighbor through the gap.]

Boy 1: Does it still work?
Girl 1: Yes, for this one.
Girl 2: Yes.

Boy 1: For the Juicy Question, I think that should be it. Like if we block out a lot of magnets, what happens?

[Experiment 2: They pull out several magnets in an attempt to stop the wave from traveling, but the wave continues.]

Girl 1: You have to cover both of the sides [of the pendulum bob] because the magnet follows the metal and the metal follows the magnet.

Girl 2: Like how do we prevent them from moving? How could we prevent them from? I mean like if you use these, how do we prevent them from moving?

[They decide to brainstorm their Juicy Questions but quickly resume speculating about how to block the effects of magnets.]

Girl 2: What would happen if the magnet wasn’t there? I mean like then it wouldn’t move.

Girl 1: It wouldn’t move at all. Because then just the metal would be there.

Girl 3: But what if a tiny piece of string was connecting them? Would it still move?

Girl 1: Yes.

Boy 1: It matters on how much it weighs.

Chaperone: What you need to put, to stop the magnet, see the paper, you put the paper [cards in between the pendulums] and it is still moving.

[Experiment 3: They try putting their Juicy Question reminder cards between the magnets and find that the pendulums still move.]

Boy 1: The magnet. Oh wait, I just, I learned this somewhere. I think magnetism can go through paper.

Boy 2: Yeah.

[Experiment 4: They again try placing their cards between the magnets.]

Girl 2: So it still works, even with the paper.

Boy 1: Get all the papers—
Girl 1: —together.
Boy 1: Now we’ll move it.

[Experiment 5: They try moving a pendulum while holding the stack of cards against its magnet.]

Boy 1: It’ll still connect [magnetically to the other pendulums] because you have to cover the metal on both sides [of the pendulum].
[Experiment 6: They try covering both sides of the pendulum with their hands and the cards.]

Girl 2: OK, will it follow?
Boy 1: It follows.
Girl 3: It follows.
Boy 2: It still goes through.

This excerpt of their inquiry at the posttest exhibit shows several Linked PAs as the group tries six experiments to investigate how to stop the pendulums from affecting one another. Experiments like these are evidence of Coherent Investigations.

**Juicy Question Helps Students and Chaperones**

A closer examination of the data reveals that the significant improvements by groups in the Juicy Question condition were made by both students and chaperones. In contrast, the improvements in the Hands Off condition were made only by the students; Hands Off chaperones did not improve on any measures. Table 5 shows the mean difference scores for students and chaperones in the four conditions, with post hoc tests comparing each inquiry condition to the mediated control condition.

This result—that chaperones learned the skills of PA and IR in the Juicy Question but not the Hands Off condition—probably reflects the asymmetrical responsibilities of chaperones in the two inquiry conditions. Chaperones played an important part in the Juicy Question game by facilitating and also contributing to the various phases: brainstorming, choosing a question, and interpreting results after an experiment. By contrast, chaperones in the Hands Off game acted mainly as referees to ensure that students respected the rules and one another. This difference may simply indicate the team’s failure to create an adequate supportive role for Hands Off chaperones, or it may reflect a deeper difference inherent in the pedagogy of the two games (insofar as Hands Off is intended to support unfacilitated use of the skills).

**Participants’ Reflections on Their Experiences**

Immediately after the posttest exhibit, we interviewed the chaperone and one randomly selected student from each group. We also phoned both participants 3 weeks later for a follow-up interview.

Exit interview responses. In the exit interview, students and chaperones in all conditions reported enjoying the experience. Chaperones and students were asked to use a 5-point Likert scale to rate the experience along a set of dimensions.
## TABLE 5
Pre/Post Difference Scores for Students and Chaperones

<table>
<thead>
<tr>
<th>Inquiry Behavior</th>
<th>Measure</th>
<th>JQ (M, SD)</th>
<th>HO (M, SD)</th>
<th>ET (M, SD)</th>
<th>Pure Control (M, SD)</th>
<th>Post Hoc Test, t(90), (p &lt; .05)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Students in group</strong></td>
<td>Number of PAs</td>
<td>3.8 (20)</td>
<td>-6.0 (20)</td>
<td>-0.35 (13)</td>
<td>1.4 (12)</td>
<td>2.6 4.7</td>
</tr>
<tr>
<td></td>
<td>PAs per minute</td>
<td>-0.05 (1.8)</td>
<td>-0.54 (1.5)</td>
<td>0.01 (1.7)</td>
<td>0.14 (1.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duration of PAs (minutes)</td>
<td>0.84 (1.2)</td>
<td>1.4 (1.4)</td>
<td>0.29 (0.81)</td>
<td>0.38 (0.87)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PA2s as a percentage of PAs</td>
<td>0.05 (0.09)</td>
<td>0.07 (0.14)</td>
<td>0.01 (0.08)</td>
<td>0.02 (0.08)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of IRs</td>
<td>13 (20)</td>
<td>2.9 (25)</td>
<td>3.6 (14)</td>
<td>-0.22 (23)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IRs per minute</td>
<td>0.8 (2.0)</td>
<td>0.48 (1.4)</td>
<td>0.51 (1.8)</td>
<td>-0.16 (2.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duration of IRs (minutes)</td>
<td>0.92 (1.0)</td>
<td>1.5 (1.6)</td>
<td>0.29 (0.99)</td>
<td>0.39 (1.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR2s as a percentage of IRs</td>
<td>0.15 (0.11)</td>
<td>0.14 (0.16)</td>
<td>0.04 (0.1)</td>
<td>0.04 (0.11)</td>
<td></td>
</tr>
<tr>
<td><strong>Chaperone in group</strong></td>
<td>Number of PAs</td>
<td>3.0 (5.0)</td>
<td>-0.13 (4.5)</td>
<td>-0.67 (4.3)</td>
<td>-0.39 (2.3)</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>PAs per minute</td>
<td>0.23 (0.44)</td>
<td>-0.03 (0.52)</td>
<td>-0.1 (0.68)</td>
<td>-0.03 (0.39)</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>Duration of PAs (minutes)</td>
<td>1.6 (2.1)</td>
<td>0.68 (2.2)</td>
<td>0.6 (1.9)</td>
<td>0.48 (2.3)</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>PA2s as a percentage of PAs</td>
<td>0.08 (0.25)</td>
<td>-0.02 (0.23)</td>
<td>-0.05 (0.27)</td>
<td>0.07 (0.23)</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Number of IRs</td>
<td>2.3 (5.4)</td>
<td>-0.28 (5.1)</td>
<td>0.41 (3.3)</td>
<td>-0.59 (3.6)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>IRs per minute</td>
<td>0.15 (0.57)</td>
<td>-0.08 (0.58)</td>
<td>0.08 (0.51)</td>
<td>-0.07 (0.63)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duration of IRs (minutes)</td>
<td>1.5 (1.9)</td>
<td>0.63 (2.3)</td>
<td>0.7 (1.7)</td>
<td>0.43 (1.8)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>IR2s as a percentage of IRs</td>
<td>0.09 (0.32)</td>
<td>0.04 (0.34)</td>
<td>0.02 (0.35)</td>
<td>0.08 (0.26)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Student data were tallied and averaged on a “per group” rather than an individual student basis. The post hoc comparisons used two-tailed *t* tests. Only comparisons with results significant at the .05 level are shown. JQ = Juicy Question; ET = Exhibit Tour; HO = Hands Off; PA = Proposing Actions; PA2 = high-level Proposing Actions; IR = Interpreting Results; IR2 = high-level Interpreting Results.
Although this was not a rigorously validated scale, we used it as simple check that the Inquiry Games did not feel onerous to field trip groups or detract from the pure exhibit experience, an important aspect of successful programs in free-choice learning environments. Students responded similarly across all conditions (i.e., we found no differences in any planned comparisons), reporting that they “had fun,” found the experience “interesting,” and felt that they had “learned something.” Chaperones also had a positive experience overall, but our planned comparisons revealed a few statistically significant differences in their responses. In one result, chaperones in the Exhibit Tour condition agreed more with the statement that they had fun than chaperones in the two inquiry conditions. Another result showed that chaperones in the Juicy Question condition felt it was more difficult to manage the group than chaperones in the Hands Off condition. Perhaps both of these results reflect the added responsibility chaperones felt as inquiry facilitators. Finally, chaperones in the Exhibit Tour condition felt that they had “learned something” more than those in the Pure Control, suggesting that they attributed value to the tour given by our educator. Table 6 shows the mean response scores for chaperones and students.

**Benefits and drawbacks to the mediated conditions.** We asked chaperones and students in the three mediated conditions to describe what they liked most and least about either the game or the tour they experienced, and we coded their responses using an emergent set of codes applied by two coders. (Participants in the Pure Control condition were not asked this question.) Planned comparisons for the effects of inquiry and pedagogy were made using Fisher exact tests. Table 7 shows that although students mostly agreed with chaperones within each condition, the depictions from the groups as a whole varied across conditions.

Groups in the Juicy Question condition more often mentioned that the game helped them think, focus, collaborate, and learn inquiry. The chaperones especially seemed to appreciate the intellectual value of the Juicy Question game. For example, one chaperone said, “I liked that it made them think. They actually had to look at the exhibit and try to solve something, trying to figure something out. It was a good thinking exercise.” Students also valued this aspect of the game, with one remarking, “I liked it because it made your brain be working more [sic], testing more, and made you think how things are done, and I think it helps you get more [out] of the objects.”

Hands Off groups felt that the game helped them learn inquiry and learn to collaborate by taking turns. One chaperone described it this way:

> I thought it was a really organized approach, and there’s more of a lesson there than just “hands off.” It’s about really listening to other people and taking turns and communicating about whatever people do. It’s a good way to practice being an adult.

A typical student response was, “It gave me a chance to be able to talk and say what I thought.”
DEEPENING STUDENTS’ SCIENTIFIC INQUIRY SKILLS

TABLE 6
Participants’ Responses About Their Experiences

<table>
<thead>
<tr>
<th>Likert Scale Responses, M (SD)</th>
<th>Juicy Question (Group Inquiry)</th>
<th>Hands Off (Individual Inquiry)</th>
<th>Exhibit Tour (Control)</th>
<th>Pure Control (Control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Had fun</td>
<td>4.3 (0.10)</td>
<td>4.2 (0.10)</td>
<td>4.5 (0.10)</td>
<td>4.4 (0.10)</td>
</tr>
<tr>
<td>Interesting</td>
<td>4.4 (0.10)</td>
<td>4.2 (0.10)</td>
<td>4.5 (0.10)</td>
<td>4.5 (0.10)</td>
</tr>
<tr>
<td>Too long (R)</td>
<td>3.9 (0.17)</td>
<td>3.8 (0.17)</td>
<td>4.0 (0.17)</td>
<td>4.1 (0.17)</td>
</tr>
<tr>
<td>Others interfered (R)</td>
<td>3.9 (0.20)</td>
<td>3.8 (0.20)</td>
<td>4.0 (0.20)</td>
<td>3.8 (0.20)</td>
</tr>
<tr>
<td>Learned something</td>
<td>4.1 (0.15)</td>
<td>4.0 (0.15)</td>
<td>4.2 (0.15)</td>
<td>4.1 (0.15)</td>
</tr>
<tr>
<td>Felt put on the spot (R)</td>
<td>3.8 (0.20)</td>
<td>4.0 (0.20)</td>
<td>3.8 (0.20)</td>
<td>3.9 (0.20)</td>
</tr>
<tr>
<td>Inquiry Game helped us</td>
<td>3.7 (1.20)</td>
<td>3.8 (1.10)</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Chaperones

| Had fun¹                     | 4.1 (0.10)                    | 4.1 (0.10)                    | 4.2 (0.10)             | 4.0 (0.10)             |
| Interesting                  | 4.5 (0.10)                    | 4.3 (0.10)                    | 4.6 (0.10)             | 4.4 (0.10)             |
| Too long (R)                 | 4.0 (0.10)                    | 3.6 (0.10)                    | 4.0 (0.10)             | 4.2 (0.10)             |
| Hard to manage group (R)²    | 3.8 (0.18)                    | 4.4 (0.17)                    | 4.2 (0.17)             | 4.2 (0.17)             |
| Learned something³           | 4.0 (0.13)                    | 3.9 (0.13)                    | 4.2 (0.13)             | 3.7 (0.13)             |
| Want more participation      | 3.6 (0.19)                    | 3.5 (0.18)                    | 3.5 (0.18)             | 3.2 (0.18)             |
| Inquiry Game helped us       | 3.7 (1.3)                     | 3.7 (1.3)                     | —                      | —                      |

Note. (R) = item was reverse scored.

¹Effect of inquiry, such that Exhibit Tour outperforms Juicy Question and Hands Off, F(1, 179) = 4.3, p < .05.
²Effect of pedagogy, such that Hands Off outperforms Juicy Question, F(1, 167) = 7.4, p < .01.
³Effect of prior mediation, such that Exhibit Tour outperforms Pure Control, F(1, 179) = 5.8, p < .06.

Finally, the Exhibit Tour participants most enjoyed learning about how the exhibits were developed, hearing about the science content, spending time with a good teacher, and simply using the exhibits. For example, one chaperone valued hearing about the development of the exhibits, saying,

I’m not a very scientific person, so the stories behind it and what we were supposed to learn were really good for me. Also learning about the people making them, and what was in their brains, I learned a lot from that. One other thing—I liked how she let the kids play with it first, then told the scientific process, then play again, then a story about the inventor.

Another chaperone expressed pleasure in learning science content: “The Bernoulli effect was fascinating. I never knew that was how airplanes flew.” Fewer students than chaperones expressed positive feelings about the stories and content in the tour. Still, one student said, “I liked to learn about how the exhibits worked and how they are made.” Another student focused on the science content:
### TABLE 7
Student and Chaperone Responses to the Mediated Conditions in the Posttest Interview

<table>
<thead>
<tr>
<th>Response</th>
<th>Students</th>
<th>Fisher p</th>
<th></th>
<th></th>
<th></th>
<th>Fisher p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Juicy</td>
<td>Hands</td>
<td>Exhibit</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Question</td>
<td>Off Tour</td>
<td>Pedagogy</td>
<td>Question</td>
</tr>
<tr>
<td>Liked most</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helped us focus our activity</td>
<td>61%</td>
<td>15%</td>
<td>7%</td>
<td>.01</td>
<td>.01</td>
<td>59%</td>
</tr>
<tr>
<td>Helped us think</td>
<td>15%</td>
<td>15%</td>
<td>0%</td>
<td>.01</td>
<td></td>
<td>59%</td>
</tr>
<tr>
<td>Helped us collaborate</td>
<td>22%</td>
<td>39%</td>
<td>0%</td>
<td>.01</td>
<td></td>
<td>17%</td>
</tr>
<tr>
<td>Helped us learn inquiry</td>
<td>20%</td>
<td>20%</td>
<td>2%</td>
<td>.01</td>
<td>.01</td>
<td>7%</td>
</tr>
<tr>
<td>Helped us take turns</td>
<td>9%</td>
<td>41%</td>
<td>0%</td>
<td>.01</td>
<td></td>
<td>4%</td>
</tr>
<tr>
<td>Enjoyed exhibit story</td>
<td>0%</td>
<td>0%</td>
<td>28%</td>
<td>.01</td>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>Enjoyed science content</td>
<td>0%</td>
<td>0%</td>
<td>24%</td>
<td>.01</td>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>Enjoyed good teacher</td>
<td>0%</td>
<td>0%</td>
<td>13%</td>
<td>.01</td>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>Enjoyed time using exhibit</td>
<td>2%</td>
<td>2%</td>
<td>13%</td>
<td>.05</td>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>No positive comments</td>
<td>7%</td>
<td>4%</td>
<td>20%</td>
<td>.05</td>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>Liked least</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard to form or choose questions</td>
<td>33%</td>
<td>4%</td>
<td>0%</td>
<td>.01</td>
<td>.01</td>
<td>41%</td>
</tr>
<tr>
<td>Hard to stop for someone else</td>
<td>2%</td>
<td>44%</td>
<td>4%</td>
<td>.01</td>
<td>.01</td>
<td>17%</td>
</tr>
<tr>
<td>Seemed silly or unnecessary</td>
<td>2%</td>
<td>2%</td>
<td>0%</td>
<td>.01</td>
<td>.01</td>
<td>0%</td>
</tr>
<tr>
<td>No negative comments</td>
<td>39%</td>
<td>39%</td>
<td>80%</td>
<td>.01</td>
<td></td>
<td>30%</td>
</tr>
</tbody>
</table>

**Note.** Response categories are not mutually exclusive; percentages do not sum to 100%. Only categories that contained at least 25% of respondents or showed a significant effect of inquiry or pedagogy are reported. Boldface indicates the condition that outperformed the other conditions.
I liked how, especially on the Bernoulli effect, how the fast moving air had different pressure, and she had us feel the pressure over and under the ball, and I liked how she demonstrated the torque thing at the next exhibit.

In terms of the negative, each condition seemed to spawn a unique problem for field trip groups. In Juicy Question, more than one third of the groups felt it was difficult to generate or choose a question to investigate. One chaperone said, “They had a hard time picking a question. They wanted to just play, so it was hard for me to actually get them to stop and think about a question.” For Hands Off groups, more than a third mentioned the challenge of stopping for someone else, presumably when the person had called out “Hands off.” One student complained, “Sometimes if you’re having fun, and they call ‘Hands off,’ you have to stop.” Finally, a small but significant number of chaperones in the Exhibit Tour condition felt that the mediation was irrelevant. One said, “The point about the dates when something was developed, like with Bernoulli, the kids probably don’t care.”

**Using the game in the future.** We also asked students and chaperones 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 no differences between the Juicy Question and Hands Off conditions, with more than 60% of chaperones and more than 30% of students predicting that they would use it. This result stands in contrast to the finding for family groups, in which fewer Hands Off families (30%) predicted that they would use the game in the future. Some of the Hands Off families felt the game was unnecessary, apparently because they felt they already knew how to take turns well. Perhaps visitors on school field trips, particularly chaperones, value a turn-taking strategy more highly.

**Follow-up interview.** To assess the groups’ longer term use of the inquiry skills they learned, we called the chaperone and one student from each group 3 weeks after they had participated in the study and asked them to self-report whether there was anything from the study that they had used subsequently. Table 8 shows the responses of students and chaperones from each condition and the results of Fisher exact tests for the effect of inquiry.

As expected, participants in the two Inquiry Games conditions more often reported that they had applied the two taught skills (PA or IR) or processes (turn taking) at new exhibits and even outside the museum. For example, a student in the Hands Off condition spoke about using it in the museum:

Me and my friend were going to the light exhibit, the one where the light flashes, and she noticed something and she said, “Hands off,” and I said, “What?” and she noticed that there was a little light blinking at the bottom of the exhibit. And it was true. And that’s how we learned about electricity.
Table 8: Student and Chaperone Responses in Follow-Up Interview

| Question                                                                 | Students                                                                 | Chaperones                                                                 |
|--------------------------------------------------------------------------|--------------------------------------------------------------------------|---------------------------------------------------------------------------|--------------------------------------------------------------------------|
|                                                                         | Juicy Hands | 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)  | 26% | 14% | 0% | —a | .01 | 14% | 13% | 0% | —a | .05 |
|                                                                         | Exhibit topic was relevant | 2% | 0% | 23% | —a | .01 | 0% | 0% | 21% | —a | .01 |
|                                                                         | Didn’t have time | 23% | 32% | 13% | —a | .01 | 43% | 50% | 34% | —a | .01 |
|                                                                         | Hard to implement | 9% | 16% | 3% | —a | .01 | 21% | 29% | 14% | —a | .01 |
| Used anything the teacher said in your life after you left the museum that day? | I used the skills or process | 19% | 19% | 0% | —a | .01 | 11% | 17% | 0% | —a | .05 |
|                                                                         | Already do something like this | 7% | 5% | 0% | —a | .01 | 21% | 17% | 3% | —a | .05 |
|                                                                         | Exhibit topics were relevant | 2% | 3% | 26% | —a | .01 | 0% | 0% | 7% | —a | .01 |
|                                                                         | I talked about the experience | 2% | 19% | 10% | —a | .01 | 0% | 8% | 24% | —a | .01 |
|                                                                         | Didn’t come up afterward | 23% | 22% | 15% | —a | .01 | 43% | 46% | 42% | —a | .01 |
| Remembered something about the exhibits?                                 | Remembered pretest exhibit | 91% | 97% | 82% | 95% | .05 | 82% | 79% | 76% | 86% | .05 |
|                                                                         | Remembered second exhibit | 100% | 100% | 97% | 95% | .05 | 89% | 96% | 97% | 93% | .05 |
|                                                                         | Remembered third exhibit | 77% | 68% | 79% | 63% | .05 | 50% | 63% | 86% | 61% | .01 |
|                                                                         | Remembered posttest exhibit | 85% | 84% | 71% | 80% | .05 | 82% | 71% | 72% | 64% | .05 |

Note. Response categories are not mutually exclusive; percentages do not sum to 100%. Only categories that contained at least 25% of respondents or showed a significant effect of inquiry or pedagogy are reported. Fisher exact tests were used to assess the effect of inquiry (Juicy Question + Hands Off > Exhibit Tour). There were no significant effects of pedagogy. PA = Proposing Actions; IR = Interpreting Results; a = Not asked. Boldface indicates the condition that outperformed the other conditions.
A student who had learned Juicy Question elaborated on how he had used the game in school:

We were doing a project for science. We took two cups and put a string through it so it was connected. You know when you pull a string it’s straight? And when it’s straight you can talk through the cup. After we tested it, we thought there is a certain length you need to make it work, like if it’s too long it wouldn’t work. It went almost the whole length of the school yard. First we played with [the cup and string], then we thought of the question, like how long the string was, and it didn’t work. It has to be a little shorter than the length of the school yard.

In contrast, students in the Exhibit Tour condition focused on connecting the science content to something else in the museum or at school. For example, one student made a connection between the magnets in the posttest exhibit and current work in the classroom: “We are learning about positive and negative magnets in school and talking about things that attract and repel.”

In short, the interview data suggested the following:

1. Students in all three mediated conditions enjoyed the experience they received, but what they liked and disliked varied. Compared to controls, field trip groups who played the Inquiry Games particularly appreciated the way the game helped them to learn inquiry and collaborate and (for Juicy Question) to focus their activity; at the same time, some groups (in Hands Off) disliked having to wait for others to finish and (for Juicy Question) negotiating to choose a question to investigate.
2. Three weeks after the experience, there were no differences across condition in the number of participants using what they had learned. However, there were significant differences in the kinds of knowledge participants reported using after the study. Most frequently, students and chaperones in the inquiry conditions applied the skills, whereas Exhibit Tour students (but not chaperones) applied the exhibit content.
3. There were two significant differences across conditions in participants’ overall memories of the experience as indicated by the number of exhibits they described having used. Students in the two inquiry conditions remembered the first exhibit better than the controls, and chaperones in the Exhibit Tour condition were better at remembering the third exhibit used. Perhaps the tour of the third exhibit in which the educator talked about balance and counterweight was particularly memorable.

Science Content Correctness

Although we did not design the Inquiry Games to help students and chaperones arrive at any particular scientific content, we endeavored to determine whether
improving their inquiry process might plausibly lead to conceptual understanding in whatever area of science they were exploring. To that end, we categorized all sophisticated interpretations of results (IR2s) as scientifically correct, incorrect, or uncodeable. We had hoped to analyze the pre/post differences in the percentage of correct IR2s made by groups in the four conditions. Unfortunately, groups in all conditions made too few IR2s in the pretest (less than two on average) to warrant categorization as correct or incorrect, so only posttest IR2s were coded. Furthermore, groups in the two control conditions did not make enough IR2s, even in the posttest, to justify any kind of comparison (see Table 9). Consequently, our analysis examined only the IR2s made by groups in the Juicy Question and Hands Off conditions.

Figure 13 shows that nearly all of the student groups in the inquiry conditions who made high-level interpretations (IR2s) in the posttest were scientifically correct in most of their assertions. In fact, almost half of the groups (41 of 86) made canonically correct interpretations every time. Recall that groups in both

<table>
<thead>
<tr>
<th>Condition</th>
<th>Total</th>
<th>Mean per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juicy Question</td>
<td>399</td>
<td>8.7</td>
</tr>
<tr>
<td>Hands Off</td>
<td>224</td>
<td>4.9</td>
</tr>
<tr>
<td>Exhibit Tour</td>
<td>86</td>
<td>1.9</td>
</tr>
<tr>
<td>Pure Control</td>
<td>100</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Note. IR2 = high-level Interpreting Results.

FIGURE 13  Number of groups in the two inquiry conditions making correct interpretations (correct IR2s) as a percentage of their total interpretations (codeable IR2s). IR2 = high-level Interpreting Results.
inquiry conditions increased the number of IR2s significantly more than controls. Together, these results show that most inquiry groups displayed correct higher level interpretations after instruction.

**DISCUSSION**

In terms of the overall research questions addressed by this study, we found the following:

- *Learning inquiry is possible.* Within a museum laboratory setting, it is indeed possible to teach field trip students skills that deepen their group inquiry, even when assessed at a totally novel interactive exhibit and without museum staff present. In particular, the two skills of publicly proposing actions and interpreting results can serve as effective metacognitive skills that bracket and extend the more typical behaviors of exploration and experimentation that students often engage in.

- *Structured activity is better than spontaneous.* Semistructured activities such as the Juicy Question game that allow groups to ask and answer their own questions but do so in a clearly defined process seem particularly effective at increasing the use of the targeted skills as well as highly desirable skills that were not explicitly taught (viz., investigatory coherence, level of PAs and IRs, and verbal collaboration).

- *Prior preparation is not needed.* Both chaperones and students can extend their inquiry by engaging in these activities. It is significant that such activities can be successfully taught on site, without the need for prior classroom preparation, and are perceived by both chaperones and students as worthwhile and enjoyable.

- *Transfer sometimes occurs.* There is evidence of occasional spontaneous transfer by a minority of students to settings beyond the research study, such as the rest of the museum visit and daily life beyond. In the case of the Inquiry Games, such transfer focuses on the use of the two targeted skills or collaborative processes such as turn taking.

- *Families and field trip groups respond similarly.* In most respects, field trip students learn and practice these inquiry skills similarly to intergenerational family groups. Comparing the results from the field trip groups to the findings in a previous study of families (Gutwill & Allen, 2010a), we observed largely similar patterns in the inquiry behaviors and attitudes of the two audiences. In an interaction analysis, we found no significant interaction effects between audience and treatment with regard to the effects of inquiry.
and prior mediation on any of our measures.\textsuperscript{11} We did find interactions with regard to the effect of pedagogy on 4 of our 13 inquiry measures, but there was no clear pattern to the interactions.\textsuperscript{12} These few differences were far outweighed by the similarities across the two audiences. In summary, the Inquiry Games, especially Juicy Question, were successful at improving participants’ inquiry behaviors in both the family and field trip groups.

Study Hypotheses

More specifically, our study of field trip groups revealed that both the Juicy Question and Hands Off Inquiry Games improved inquiry behaviors at a novel exhibit compared to a control condition. The games helped groups make (a) longer, more sophisticated proposals for action and interpretations of results and (b) more collaborative interpretations. Further analysis revealed that the Juicy Question condition improved additional aspects of groups’ inquiry, such as making more interpretations and conducting more Coherent Investigations. It seems that the structured collaborative format was more effective than the spontaneous individualistic one.

In interviews, participants reported that the Juicy Question condition helped them focus and think, whereas the Hands Off condition helped them take turns. Three weeks after the experience, a small but significant fraction of participants in both inquiry conditions reported that they used the skills they had learned in other situations outside the research lab.

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

Hypothesis 1: Inquiry Games will improve inquiry behaviors. We found that groups who learned the Inquiry Games outperformed those in the mediated control condition on several measures of inquiry, thus confirming the hypothesis that lay at the crux of the study. We found a similar result in our previous study

\textsuperscript{11}To determine whether the Inquiry Games affected field trip and family groups differently, we compared the inquiry behaviors from the two studies. Using a univariate analysis of variance, we set up contrasts for the interactions between audience (family vs. field trip) and the original planned comparisons (effect of prior mediation, inquiry, and pedagogy). Statistically significant interaction effects indicated differential effects of the treatment conditions on the two audiences.

\textsuperscript{12}We found interaction effects for the following four variables: number of PAs, duration of IRs, Level 2 IRs as a percentage of IRs, and Consecutive IRs as a percentage of IRs.
of the family groups, indicating that the Inquiry Games may be useful tools for museums to use with different audiences.

It may be argued that this finding is unsurprising: We taught some inquiry skills and learners could use them. However, we believe that there are several reasons the result goes well beyond basic skill acquisition. First, the participants were posttested using their skills in the context of a novel exhibit with different science content, a more challenging task than a traditional repeat-measures test. In this sense, the project can be seen as study of near transfer and supports our earlier conjecture that skills may be more readily transferable than concepts or principles. Second, the same basic pattern of results was seen in two important aspects of inquiry that were never taught at all, namely Collaborative Explanations and Coherent Investigations. This suggests that the skills taught in the Inquiry Games may have been gateways to other skills, perhaps because of an underlying mechanism of social group coherence. More specifically, for group members who want to support and listen to each other, the activity structures of the game scaffold them in publicly stating their actions and interpretations. Consequently, group members may fall naturally into collaborative activity, building on each other’s ideas so that no member is left feeling isolated or foolish. If so, our result suggests that within-group bonds may indirectly facilitate some inquiry processes (such as coherence of investigations), just as they may undermine others (such as competitive argumentation). Third, the increase in skill use was not substantially correlated with any of the demographic variables of either students or chaperones, suggesting that the Inquiry Games worked equally well across a range of participants. This probably reinforces our original claim that schools and museums are inherently quite different activity systems and suggests that informal environments may support low-performing students in particular.

Finally, it is interesting to consider this finding in relation to debates about whether skills can be learned separately from content. We would argue that the skills taught in the activity structure of the Inquiry Games were essentially metacognitive skills, generic enough to be broadly applicable and yet leading learners to engage with the specific interactions and science content inherent in each exhibit design (e.g., pulling back different combinations of coupled pendulums to discover the system’s tendency to lock into a single phase). We believe the skills could serve as successful generic gateways because of the motivating influences of contextual factors: (a) The physical context provided intriguing phenomena worthy of deeper exploration, thus motivating learners to dig into their specific characteristics; and (b) the social context allowed learners to investigate in directions of their own choosing. At the same time, we acknowledge that the degree to which students ultimately learned canonical science content remains an open question—without an educator whose goals were to emphasize particular concepts or principles, many groups discovered content that was not immediately relevant to the school science curriculum.
Hypothesis 2: Prior mediation will enhance the experience. Based on previous research showing positive effects of mediation on visitors’ attitudes (National Research Council, 2009), we expected that groups in the mediated control condition (Exhibit Tour) would enjoy the experience more than those in the unmediated control condition (Pure Control). Moreover, we predicted that improved attitudes would lead to more engagement and thus better inquiry by the time the group reached the posttest exhibit. Results indicated little, if any, effect of prior mediation on participants’ attitudes or inquiry skills. Attitudinally speaking, the only difference we found on any measure came in the form of more chaperones in the Exhibit Tour condition than in the Pure Control feeling that they had “learned something” (students were no different in the two conditions). In terms of inquiry behaviors, only 1 of 13 assessments showed a difference: Exhibit Tour groups increased the frequency of their interpretations more than Pure Control groups. Based on these small differences, we conclude that our hypothesis for prior mediation was disconfirmed. Of course, it is important to acknowledge the limited scope of this finding: It is entirely possible that prior mediation in a content-relevant area would enhance content-based learning; it is also entirely possible that the presence of a mediator during the posttest would enhance inquiry-based learning. Neither of these was tested in the current study. Viewed in a more positive light, the results indicate that the Exhibit Tour was an effective control for the two inquiry conditions because it successfully offered mediation without overly stimulating groups or teaching them inquiry skills.

Hypothesis 3: Pedagogy will affect inquiry. We predicted that the collaborative nature of Juicy Question might make for deeper inquiry, whereas the focus of Hands Off on turn taking might make it easier for students to work together. Our results indicate that Juicy Question was more effective than Hands Off at improving a broad range of inquiry behaviors, including the number and frequency of proposed actions, the number of interpretations, and the coherence of investigations. Meanwhile, Hands Off was better on only one measure: duration of proposed actions. We found a similar pattern of superiority for Juicy Question in our study of family groups. We conclude that the Juicy Question game is a more effective educational intervention for a wider variety of science museum visitors.

Understanding Juicy Question’s Success

We were not surprised that the two inquiry conditions outperformed controls, given our extended iterative design process and the use of accepted learning principles from both school and field trip settings. However, to gain a qualitative sense of why Juicy Question outperformed Hands Off in terms of making proposals of action, interpreting results, and linking multiple investigations, we reviewed video and interview data from a total sample of 24 high-improving and low-improving
Everyone must participate. The Juicy Question game asked all students to try generating fruitful questions and interpreting results; in Hands Off, shy students could decline to call “Hands off!” and thereby avoid practicing the skills. Although our intention in creating the Hands Off game was to encourage involvement, students in larger groups would find it easy to retreat into silence. Our video review found several instances of outgoing students maintaining control of the exhibit by repeatedly calling “Hands off” to suggest a new course of action. In interviews, some students in the Hands Off condition complained about this, with one remarking, “Some people, they didn’t get to say ‘Hands off’ because some people were pretty much taking all the spotlight.” By allowing extroverted children to dominate the activity, the Hands Off game may have inadvertently excluded introverted students, leading to a smaller improvement in inquiry behaviors overall. In Juicy Question, every student was encouraged to participate in the brainstorm and discovery phases, which may have strengthened the skills of the group as a whole. Of course, students’ motivations for participating are likely to be complex and were not assessed in depth in this study; it may be that some of the quieter students would have been more verbal had they believed their participation would be factored into their school grades. Also, quieter students may nevertheless have learned a great deal from their experience; our study was limited in that the assessments focused on externalized discourse and gestures and did not address the kinds of learning-by-watching that have been shown to be effective in other settings (Rogoff, Paradise, Mejia Arauz, Correa-Chavez, & Angelillo, 2003).

Group members collaborate to choose a question. During the brainstorming phase of Juicy Question, students must work together to choose one question to investigate. This was sometimes difficult for them, as mentioned in the interview by 33% of the students and 41% of the chaperones. Video analysis suggested that some groups made this easier by brainstorming questions only until an idea piqued the group’s interest, then investigating it immediately rather than completing the brainstorm process. Whether groups used the standard brainstorm procedure or the abridged version, the process probably yielded several positive outcomes. First, filtering out lower quality questions in favor of “juicier” ones probably led to more interesting experiments worthy of deeper discussion and interpretation. In contrast, Hands Off allowed each student to pursue a stated question no matter how trivial or simplistic it might have been. One student complained of this problem in the interview, saying, “Sometimes people called
“Hands off” for pointless reasons.” Second, agreeing on the question to pursue could enhance Juicy Question group members’ interest and investment in their experiments, leading to more discussion, new questions, and more coherent sets of investigations. Hands Off may unintentionally encourage abrupt shifts in the topics driving investigations as one student after another announces an unrelated idea and begins investigating it. As one interviewee noted, “Some kids don’t like you to call ‘Hands off’ because it distracts them from their work.” Finally, the collaborative nature of the Juicy Question format may have produced greater feelings of camaraderie, as reflected in videos in which group members used terms connoting group effort, like “We’re almost there” and “We’ve answered the question! We’re good!” and “Look, we got it!” The emotional comfort that comes from amity may encourage greater participation in group experimentation and discussion.

**Juicy Question includes an explicit interpretation phase.** The final stage of the Juicy Question game asks group members to pause and reflect on the discoveries they made during their investigation. This encourages all participants to practice the metacognitive skill of interpretation. In the Hands Off game, calling “Hands off” to state a discovery is optional. In fact, the “question” phase of Juicy Question almost demands an “answer” phase, whereas the “plan” activity for Hands Off can be viewed as independent of the “discovery” activity—plans do not necessarily lead to discoveries. This may explain why the Juicy Question groups increased the number of times they interpreted results so much more than the Hands Off groups. Furthermore, our video review found that the interpretation phase of Juicy Question often sparked new, related questions, thus producing longer sets of linked experiments.

**Chaperones make an important contribution.** The overall success of Juicy Question was due in part to chaperones improving their own inquiry behaviors compared to chaperones in the Exhibit Tour Control condition. (In contrast, chaperones in the Hands Off condition performed no better than those in the Exhibit Tour condition.) By placing chaperones in the role of facilitator, the Juicy Question game somehow engaged them more deeply in the inquiry process.

To learn more about the nature of chaperone facilitation in the Juicy Question condition, we turned to our educator ratings, a set of dichotomous ratings our educators made to describe the chaperone’s performance during the treatment phase of the experiment. For example, our educators rated chaperones as “skilled” or “unskilled” facilitators, as “engaged” or “not engaged” in the spirit of the game, as sticking to the structure of the game “exactly” or “not exactly.” These ratings were neither validated nor checked for reliability; their original intended use was primarily to help the team identify interesting cases along a spectrum of chaperone performance. Recognizing this limitation, we nevertheless explored these data
following the main quantitative analysis phase to see whether they might shed additional light on the role of the chaperones in scaffolding group processes. Specifically, we used the ratings as independent variables to see whether chaperone performance during treatment was related to group inquiry performance in the posttest. Of the educator ratings, the facilitator rating showed a significant and surprising pattern of differences on four of our dependent measures: Juicy Question groups with unskilled chaperone facilitators made more IRs, IRs per minute, Linked PAs, and Consecutive IRs than Juicy Question groups with skilled chaperone facilitators. Moreover, the unskilled chaperones themselves made significantly more IRs per minute than the skilled chaperones, as did their respective students. (Linked PA and Consecutive IR scores were assigned to entire groups, so we could not disaggregate chaperone from student performance on those measures.) These findings suggest that in groups with unskilled chaperones, both the chaperones and the students were contributing more than in groups with skilled chaperone facilitators. Again, the educator ratings were not validated, but the recurrence of the same pattern of results across four of our key dependent measures prompted us to report it here.

If accurate, these results suggest that when chaperones are less comfortable or adept at facilitating the Juicy Question game, they tend to slip into the students’ role, co-investigating exhibit phenomena along with the students. Meanwhile, students in such groups tend to contribute more, perhaps because their chaperones are effectively modeling the inquiry skills with adult-level expertise, or perhaps because they step into a perceived leadership vacuum, or perhaps because they benefited earlier from witnessing their chaperone receive more extensive coaching from the staff educator who perceived them as unskilled. Whatever the reason, we see it as a value of the Juicy Question game that it seems to work particularly well in advancing the inquiry of groups with unskilled chaperones.

If true, these results also have implications for the theoretical framework underlying our experiment. We designed our inquiry conditions with the assumption of nested zones of proximal development, in which museum educators help parent chaperones learn to facilitate, and chaperones in turn help students engage in game-based inquiry. We assumed that for each level, the more experienced person would create a zone of proximal development by taking on the most challenging tasks and structuring them for others. We expected that when the educator departed in the posttest, chaperones would create a zone of proximal development for students by facilitating the game. Some facilitators (the skilled ones) probably did establish such a zone, but the largest gains apparently came from groups in which chaperones acted more like students than facilitators. This implies that a flat rather than nested zone of proximal development treatment structure may be more effective; that is, perhaps Juicy Question would be even more successful if we did not ask chaperones to facilitate the game. Further research is needed to shed light on these questions.
Limitations of the Study

As with the study of family groups previously mentioned, the field trip study has several limitations that may narrow the applicability of the results. First, even though the groups were left alone to use the last (posttest) exhibit, before leaving the educator did explicitly ask those 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 groups to use a final exhibit “as they normally would” after we had given them game cards and spent 20 min teaching them an exhibit game—seemed disingenuous; also, it would have led to a confounding, because we would not have known whether the groups 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 the field trip groups could play the games at the posttest exhibit. More important, even in ideal circumstances, simply telling someone to engage in a particular learning process does not mean that they will have the motivation or ability to comply. Barbara White, an experienced inquiry researcher and project advisor, commented, “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” (personal communication, February 14, 2005). Similarly, our experiences with visitors during the early development of the Inquiry Games made it clear that giving a broad directive to play the game was frequently ignored, especially if the visitor group did not find the game memorable or enjoyable.

Second, we coached and assessed inquiry behaviors at only one type of science museum exhibit, which we characterized as an interactive, open-ended, multioption, and multiuser exhibit. The results of the study may be less applicable to so-called discovery-based 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 field trip groups 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 field trip groups had the exhibits to themselves, could spend as long as they wanted, could hear one another clearly, and endured few

---

13In 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 field trip group inquiry arguably pertains to three distinct exhibits.
distractions while using each exhibit. How well would the Juicy Question game function on the buzzing museum floor, where students 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.

Future Implementation

To learn more about how the Juicy Question game fares in the exciting environment of a real science museum, we are currently adapting the game for use with families and field trip groups on the museum floor and formatively evaluating the results of our endeavors. Several issues have already emerged, unique to each audience.

For field trip groups, implementation challenges include the following:

- **Large student groups.** In our lab study, we limited group size to seven students. On real field trips to the Exploratorium, groups are supposed to have no more than 10 students per chaperone, but sometimes the ratio is higher. Students in large groups have more difficulty hearing one another, getting an opportunity to manipulate the exhibit, and having the time to participate in the conversation.

- **Learning the game takes time.** Even with the games limited to two skills, the time required to teach Juicy Question at an exhibit sometimes exceeds the attention span of students and their chaperone, especially for large groups.

- **Increased distractions.** Unlike the laboratory, the museum floor is full of interesting exhibits, people, and sounds, all of which can distract students from learning the Juicy Question game.

- **Chaperone priorities.** On the museum floor, chaperones often feel the need to focus their energies on keeping track of students rather than playing the game. Perhaps the game will be more effective if students are asked to facilitate the game for themselves.

To deal with these challenges, we are exploring various techniques, such as teaching students and chaperones about the game off the floor and then applying it to a real exhibit as a second step.14

**ACKNOWLEDGMENTS**

We gratefully acknowledge the creative and analytical contributions of all members of the GIVE (Group Inquiry by Visitors at Exhibits) team, particularly

---

14Updates on the project’s progress can be found at www.exploratorium.edu/partner/give.
Ryan Ames, Mark Boccuzzi, Liana Crouch, Sarah Elovich, Beth Gardner, Malia Jackson, Adam Klinger, Nerissa Kuebrich, Suzy Loper, Anne Richardson, Lisa Sindorf, and Erin Wilson. We also wish to thank Linda Deacon, John Frederiksen, Kathleen McLean, Michael Ranney, Barbara White, 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. This material is based upon work supported by the National Science Foundation under Grant 0411826 and work supported by the Foundation done while Sue Allen was working at the Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are our own and do not necessarily reflect the views of the National Science Foundation.

REFERENCES


