Participant Reactivity in Museum Research: The Effect of Cueing Visitors at an Interactive Exhibit

Scott A. Pattison and Todd Shagott

Oregon Museum of Science and Industry, Portland, Oregon, USA

Article citation: Scott A. Pattison & Todd Shagott. (2015). Participant reactivity in museum research: The effect of cueing visitors at an interactive exhibit, *Visitor Studies, 18*(2), 214-232, DOI: 10.1080/10645578.2015.1079103

About the Authors

Scott A. Pattison, Ph.D., is a researcher and evaluator at the Oregon Museum of Science and Industry and the Institute for Learning Innovation. He studies education, learning, and interest development in free-choice and out-of-school environments, including museums, science centers, and everyday settings. Address correspondence to Scott A. Pattison, Oregon Museum of Science and Industry, Portland, OR 97214. E-mail: spattison@omsi.edu.

Todd Shagott, Ph.D., is a research assistant at the Oregon Museum of Science and Industry. His work is currently focused on exhibit evaluation and visitor interactions. E-mail: tshagott@omsi.edu.

Acknowledgments

We are grateful to all the OMSI staff members and *Design Zone* project team members who supported and contributed to this work, including Marcie Benne, Kendall Bartholomew, Karyn Bertschi, Mary Kay Cunningham, Scott Ewing, Angela Fray, Cecilia Garibay, Marilyn

Johnson, Jan Mokros, and Andee Rubin. Thanks also to Marcie Benne, Karyn Bertschi, Chris Cardiel, Katura Reynolds, Debbie Siegel, Barry Walther, and two anonymous reviewers for their thoughtful feedback on initial drafts. This material is based upon work supported by the National Science Foundation under grant number DRL-0714634. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Abstract

Although studies in a variety of settings suggest that participant reactions to the research context can threaten the validity and generalizability of study findings, there have been almost no investigations of participant reactivity in museums. In this experimental study, we compared the behaviors and learning outcomes of visitors at two versions of an interactive mathematics exhibit who had either been actively recruited by a data collector or passively recruited using posted signage. We assessed the amount of time visitors spent at the exhibit, the number of mathematical exhibit behaviors they engaged in, and the level at which they described the mathematical relationships in the exhibit after the interaction. The results indicate that actively recruiting visitors was associated with increased engagement times and number of mathematical exhibit behaviors and that recruitment method moderated the relation between exhibit version and learning outcomes. These findings emphasize the importance of carefully considering recruitment decisions in museum research and evaluation.

Keywords: participant reactivity, research methodology, recruitment, validity, experimental design, interactive exhibits

Participant Reactivity in Museum Research: The Effect of Cueing Visitors at an Interactive Exhibit

Researchers studying behavior and learning in museums and science centers face unique methodological challenges (Allen et al., 2007; National Research Council, 2009). Because investigators in these contexts are often interested in naturalistic behaviors, recruitment and sampling decisions in particular can be critical, involving trade-offs between internal and external validity (National Research Council, 2009; Shadish, Cook, & Campbell, 2001). For example, recruiting visitors from a museum lobby and leading them to a separate room designed specifically for the research study may allow for easy random assignment between experimental conditions, thus supporting internal validity, but may also change participant behavior to the extent that results are difficult to generalize to the museum floor, thus limiting external validity. Research outside the museum field strongly suggests that participant reactivity to the research context can have a powerful effect on study findings and that “alert, aware participants are actively seeking cues in the research setting to inform what they are expected to do in order to present themselves in a favorable light” (Brewer & Crano, 2013, p. 17). Design decisions, therefore, that influence how participants react, such as recruitment strategies, are potentially critical to study outcomes.

Little is known about the importance of participant reactivity in museum and visitor studies research and there are few studies to guide the methodological decisions of evaluators and researchers in these settings. To address these gaps, we conducted an experimental study to compare the behaviors and learning outcomes of exhibit visitors who had either been passively recruited, using posted signage, or actively recruited by a data collector to engage with an interactive mathematics exhibit. The goal of the study was to assess the impact of active recruitment on participant reactivity in the context of museum research and provide empirical guidance for evaluators and researchers making decisions about recruitment and sampling.

Participant Reactivity

Participant reactivity is a broad term describing the potential for research participants to alter their behaviors when being observed or when involved in a research study (Brewer & Crano, 2013; Shadish et al., 2001). These behavior changes have been associated with participant apprehension and anxiety, reactions to unconscious cues provided by researchers, participants' tendency to try to guess the purpose of the research and alter their behavior accordingly, and individuals' natural inclination to behave in a socially desirable or appropriate manner (Brewer & Crano, 2013; Chiesa & Hobbs, 2008; Shadish et al., 2001). The term reactivity is often used interchangeably with the *Hawthorne Effect* (cf., Chiesa & Hobbs, 2008) and encompasses a variety of more specific concepts, such as demand characteristics, experimenter expectancies, evaluation apprehension, and social desirability bias (Brewer & Crano, 2013; Groves, Fowler, Couper, Lepkowski, & Singer, 2009).

In experimental studies, participant reactivity can threaten either construct or internal validity, confounding the effects of the treatment or weakening evidence of a causal relation between independent and dependent variables (Brewer & Crano, 2013). In other words, participant reactions to the research context can become part of the treatment, as when subjects react to the novelty of the situation, or participant reactions can differ between conditions and offer an alternative explanation for study results, such as when individuals feel more pressure to perform in the treatment compared to the control condition. More generally, participant reactivity is a potential threat to the external validity and generalizability of any type of research study when the behaviors of research participants are substantially different within the research context compared to what they might be in a more naturalistic setting (Shadish et al., 2001).

Although scholars have long identified reactivity as an important issue in social, behavioral, and psychological research, findings from investigations that have attempted to directly measure participant reactivity are mixed. In addition to early research on the Hawthorne Effect (Landsberger, 1958), recent studies from a variety of fields have demonstrated evidence of reactivity (e.g., Haidet, Tate, Divirgilio-Thomas, Kolanowski, & Happ, 2009; Spiers, Costantino, & Faucett, 2000). For example, Eckmanns, Bessert, Behnke, Gastmeier, and Rüden (2006) found that the use of antiseptic hand rub increased 55% when participants were aware that hand sanitation compliance was being monitored. Similarly, in research comparing different approaches to assessing consumer food preferences, observations by the experimenter appeared to influence the amount of chocolate participants took home (Weiss, O'Mahony, & Wichchukit, 2010). In a quasi-experimental community development study focused on the effectiveness of an indoor air pollution intervention in Africa, Barnes (2010) used post-study interviews with participants to identify issues of participant reactivity, including the perceived need to impress researchers,that made study findings impossible to interpret.

Other research, however, has found no evidence of the influence of participant reactivity on study findings. Based on a study of college students taking an introductory psychology class, Haddad, Nation, and Williams (1975) argued that academic performance and retention were not influenced by students being explicitly told that they were part of an experiment. Similarly, an investigation of physician case reviews (Fernald, Coombs, DeAlleaume, West, & Parnes, 2012) suggested that doctors who knew that they were going to undergo an intense case review process did not provide better care than those who were not involved in the reviews. In the context of family research, Jacob and colleagues (1994) found that families behaved no differently when they knew that audio recordings were being made compared to when they were unaware of the recordings. The authors speculated that family patterns are too strong and family duties too important to be influenced by observation.

One explanation for these mixed findings is that the degree of participant reactivity is moderated by the research context (Maisto, Clifford, & Davis, 2007). We know of only one study that has directly tested this hypothesis. Comparing staff hand hygiene at three healthcare units, Kohli and colleagues (Kohli et al., 2009) found that known observers temporarily influenced hand hygiene compliance in the units that already had high rates of compliance but not in the unit with low compliance rates. The authors speculated that the high compliance units may have had more at stake compared to other units and thus may have been more likely to alter their behavior when known observers were present.

Reactivity in Museum Research

Research outside museums demonstrates that participant reactivity can have important consequences for research results, although the extent to which reactivity is an issue may depend on the research context. These results highlight the need for museum researchers to understand how and when participant reactivity may influence study results. Unfortunately, there has been little research on recruitment, sampling, or reactivity in museums. Two notable exceptions are studies by Serrell (2000) and Gutwill (2003).

Serrell (2000) compared the total time cued visitor groups spent in an exhibition compared to uncued visitor groups across 13 different exhibition evaluation studies. Cued visitors were actively recruited by a data collector before entering the exhibition and told that they would be asked questions at the end of their experience, whereas uncued visitors were covertly observed without being recruited and may or may not have been aware that they were part of an evaluation study.

For 12 of the 13 exhibitions, cued visitors spent more time than uncued visitors and for 10 of these exhibitions these differences were statistically significant. Increases in average total time compared to uncued visitors ranged between 10% and 100%. The author speculated that cueing may influence engagement time by providing a physical and conceptual orientation to the exhibition, thus increasing attention, or by providing extrinsic motivation to spend more time in the exhibition. She recommended that “evaluators reporting on information gathered under cued conditions should always state what their assumptions are about cued data and whether the data can be generalized to uncued visitors” (Serrell, 2000, p. 4).

Although Gutwill (2003) did not directly investigate participant reactivity, his seminal methodological study did indicate that visitors are aware of being observed, even when they are passively recruited through signage rather than actively recruited by a data collector. Gutwill investigated the effectiveness of a posted-sign method of informed consent for notifying museum visitors that interactions at an exhibit component were being videotaped. In the study, several exhibits were placed together inside a stanchioned research area, with only one of the exhibits videotaped at any given time. Bilingual (Spanish/English) signage notifying visitors that videotaping was in progress was placed at the entrance to the museum, the entrance to the exhibit area under study, and on the videotaped exhibit component. Of the 200 adult visitors interviewed as they exited the area, 197 (99%) reported that they knew they had been videotaped and the three that reported not being aware of the research indicated they were not bothered by it.

Research Questions

Together, the two studies reviewed above provide initial evidence that visitors will change their behavior when they are aware of being observed and that this awareness can be triggered by passive recruitment, using posted signage, as well as active recruitment by a data collector. However, the research leaves many questions unanswered. In particular, it is not clear the extent to which recruitment and participant reactivity might affect other types of behaviors or outcomes beyond the total time visitors spend in an exhibition. There is also a need to compare other types of recruitment methods beyond convert observations, such as actively recruiting visitors and asking them to participate in a study and passively recruiting visitors using posted signage.

The purpose of the current study was to address these outstanding questions and extend the limited literature on recruitment and participant reactivity in museums. The study was designed to address two research questions: (1) How, if at all, do the behaviors and learning outcomes of museum visitors at an interactive exhibit change based on whether the visitors are actively recruited by a data collector or passively recruited through signage? and (2) How, if at all, do these approaches to recruitment influence the results of a study comparing the impact of two different versions of the exhibit?

Method

To address these questions, we conducted an experimental study to explore the impact of recruitment on visitor learning behaviors at an interactive exhibit in the Oregon Museum of Science and Industry (OMSI), Portland, Oregon. Specifically, we compared the behaviors and learning outcomes of visitor groups that were actively recruited by a data collector with the behaviors and outcomes of groups that were passively recruited using posted signage. The data and findings reported in this article were part of a larger experimental study to investigate the impact of a supplemental computer guide on visitor learning behaviors and outcomes at an interactive mathematics exhibit (Pattison, Ewing, & Frey, 2012). Because pilot observations during the initial study indicated that the effect of the computer guide might depend on how visitors were recruited, we included recruitment method as a separate factor in the experimental design.

Participants

The target population for the study was families visiting OMSI who chose to enter the *Design Zone* exhibition, described below. Families were defined as visitor groups with at least one adult 18 years or older and one child between the ages of 7 and 17 years, with age estimated by the data collector. The focus on families with children 7 years of age or older was based on the intended audience of the exhibition. In total, 128 families, representing 387 individuals (176 adults and 211 children), were recruited for the study. The number of family members per group ranged from two to nine, with an average of approximately three members per group. The average child age was 9.8 years and the vast majority of groups (87.5%) included either one or two children. Adult and child participant groups were approximately evenly split by gender. The majority (64.1%) of adults who participated in the interviews reported that they had not visited the museum in the last six months and only 14.8% indicated that they had previously visited the exhibition.

The target sample size of 128 families was determined in advance based on a power analysis. Assuming an alpha level 0.05, the sample size provided acceptable power (0.80) to reliably detect medium effect sizes (f = 0.25) for both main effects and interactions (Faul, Erdfelder, Lang, & Buchner, 2007).

Procedures

The study followed a 2 x 2 factorial design, with families either actively recruited by the data collector or passively recruited through a posted sign and then assigned to interact with one of two versions of the exhibit, either with or without the computer guide. The exhibit activity, which is described in detail by Pattison and colleagues (2012), was part of a 5,000 square foot National Science Foundation-funded mathematics exhibition, *Design Zone*, focused on encouraging science center visitors to understand and use the relationships between quantities to meet creative challenges in a variety of mathematically rich contexts, such as design, engineering, art, and music. At the activity used in this study (Laser Light Show), visitors adjust the speed of two oscillating mirrors to control the pattern created by a laser reflecting off those mirrors. The exhibit includes a numerical readout of the oscillation speed of each of the mirrors to support visitors in exploring the relation between the ratio of the two mirror speeds and the resulting laser pattern. A computer guide, situated to the left of the control console, provides visitors with a more structured way to explore the exhibit by posing incrementally more complex challenges.

In the study, actively recruited participants were selected based on a systematic random sampling approach (Bernard, 2006). The data collector drew an imaginary line in front of the entrance to the *Design Zone* exhibition and approached the first family with at least one child between the ages of 7 and 17 who crossed the line. To determine eligibility, the age of children was estimated by the data collector and later confirmed during the recruitment process. Eligible families were told that the science center was testing a new exhibit activity and were offered two free general admission tickets for participating in the study. Those who agreed to participate were led back to the exhibit, informed that they would be observed and then interviewed, and told to spend as long or as little time as they wanted with the activity. During data collection, stanchions were placed around the exhibit to minimize interference from other visitors.

For passively recruited participants, a bilingual (Spanish/English) sign notifying visitors about the research was posted in front of the exhibit and all families and visitor groups were allowed to engage with the activity. The informed consent sign indicated that the exhibit was being observed for research purposes and that visitor groups might be approached for an interview prior to their interaction with the activity. Data were only collected from groups in the target population, with family member age estimated by the data collector and then confirmed during the post-interaction interviews. As in the active recruitment condition, stanchions were placed around the sides of the exhibit to minimize interference from other visitor groups. Passively recruited families were also offered an incentive of two free general admission tickets after they finished interacting with the exhibit. To ensure that both recruitment groups were equivalent in terms of willingness to complete the survey, those who did not agree to participate after they had interacted with the exhibit were excluded from the study. Based on Gutwill (2003), we assumed that the vast majority of visitor groups in the passive recruitment condition were aware that they were being observed.

Rotation of recruitment condition was structured so that each condition was equally represented by time block and day of the week. Each data collection day was divided into a morning time block (10:00 a.m. to 1:00 p.m.) and an afternoon time block (1:30 p.m. to 4:30 p.m.) across four weekend days between April 16 and May 22, 2011. If participants were actively recruited during the morning time block, then data would be collected from passively recruited participants during the afternoon. This order was reversed the following weekend. For example, if data were collected from passively recruited participants on Saturday morning, participants would be actively recruited the morning of the following Saturday.

Within each recruitment condition, families were systematically assigned to interact with one of two versions of the exhibit: (a) the current version, with the computer guide available to visitors (computer condition), and (b) a modified version, with the computer guide covered and only the explanatory label available (no computer condition). Every two family groups, the data collector alternated between the computer and non-computer conditions by removing or replacing a covering over the computer screen (for a similar design, see Gutwill, 2006). This design allowed us to not only investigate the overall impact of recruitment condition on visitor behavior and learning but also explore possible interactions between recruitment and experimental condition, as suggested by pilot observations. Sample sizes were approximately equal across the four experimental conditions: 30 families in the no computer, passively recruited condition; 34 families in the no computer, actively recruited condition; 32 families in the computer, passively recruited condition; and 32 families in the computer, actively recruited condition.

Data Collection

We collected data through naturalistic observation and post-interaction interviews in order to measure visitor engagement time, mathematical exhibit behaviors, and articulation of mathematical relationships within the exhibit. In the original study (Pattison et al., 2012), these measures served as the primary dependent variables to test the effect of the computer guide on visitor behaviors and learning outcomes. For the analyses reported in this article, these measures were used to understand how recruitment method influenced visitor behaviors and learning outcomes at two versions of the exhibit. For more details about the exhibit activity and data collection and coding procedures, see Pattison et al. (2012).

Observation. While families engaged with the exhibit, the data collector recorded the amount of time spent at the exhibit and the mathematical ratio for each laser pattern created by the families. For actively recruited participants, *engagement time* began after visitors were led to the exhibit and all of the family members had entered the stanchioned area. Timekeeping ended as soon as the last adult family member exited the area. For passively recruited participants, the observation began when one adult and one child entered the stanchioned area together, a child joined an adult family member inside the stanchioned area, or an adult joined a child family member. This protocol was designed to focus data collection on family interactions and to align with the engagement time measure for actively recruited groups. Again, timekeeping ended when the last adult in the family group left the stanchioned area. In both conditions, groups that engaged in the activity for less than 30 seconds or did not create at least one laser pattern were excluded from the analysis in order to focus on visitors that engaged more deeply with the exhibit and were more likely to understand the main messages of the activity. During the observations, it was not uncommon for one or two family members to leave for a short time and then return.

The data collector also recorded the mathematical ratio, as indicated by the exhibit displays, for each laser pattern created by the family. Ratios were only recorded when families paused for at least two seconds on a pattern. For the analysis, *mathematical exhibit behaviors* were operationalized as the total number of unique laser patterns visitors created for the seven most common mathematical ratios (1:1, 2:1, 1:2, 3:2, 2:3, 3:1, and 1:3). To be counted as unique, a pattern could match the simplified ratio of a previously created pattern as long as the numerator and denominator of the patterns were distinct (e.g., 10:20, 20:40, and 15:30 would each be counted as unique patterns). The subset of seven ratios was chosen based on the assumptions that families were more likely to intentionally create patterns with these ratios and to understand their underlying mathematics. The results of the analyses reported below were identical when a broader range of ratios was included.

To check reliability of the mathematical exhibit behaviors measure, two data collectors were present during the first two days of testing. Inter-rater agreement for the 33 observations collected during this time (25.8% of all observations) was nearly perfect (Pearson's correlation, $r = 0.978$). Only one data collector was present during the remaining testing days and only data from the primary data collector were used in the analyses described below.

Interviews. After families were finished with the exhibit, the data collector asked one self-selected adult two open-ended questions about their perception of the main messages and learning outcomes of the exhibit: (a) What would you tell another visitor this exhibit is about? and (b) What do you hope the child/children in your group learned from this exhibit? Responses to these questions were coded together to measure the extent to which the adult interviewees were able to articulate the mathematical relationships in the exhibit.

The interview coding scheme was developed inductively by the first author, in consultation with the exhibit developers, and then applied to all 128 responses. A second researcher who had not participated in the development of the codes then used the initial coding scheme to categorize the first 58 interview responses (45.3% of the total). Inter-rater agreement for these responses was 70.7%. The researchers then met, resolved all disagreements, and updated the coding scheme. Subsequently, the second researcher coded the remaining 70 interview responses (54.7% of the total). Inter-rater agreement for these responses was 87.1% and all disagreements were again resolved between the two researchers.

The final coding scheme defined five levels of successively more sophisticated articulation of the mathematical relationships in the exhibit: (1) no mathematical content, (2) identifying mathematical elements of the exhibit, (3) relating two mathematical elements, (4) relating three mathematical elements, and (5) generalizing relationships. If interviewees mentioned at least one of the mathematical elements in the exhibit, such as the speed of the oscillating mirrors or the shape of the laser pattern, their responses were coded as level 2. If they described a mathematical connection between these elements, their responses were coded as level 3 or 4, depending on the number of elements described. Responses that articulated generalized relationships between mathematical elements, such as how the laser always creates a figure eight when one of the mirrors is oscillating at twice the speed of the other mirror, were coded as level 5.

During the interviews, family demographic information was also collected, including family size, adult and child gender, age of all children, frequency with which the adult interviewee had visited the science center in the last six months, and whether or not the adult had visited the exhibition before. These variables were intended to check the equivalence of visitor groups across the conditions and to serve as control variables in the multivariate analyses, as appropriate.

Data Analysis

Data were analyzed through descriptive and inferential statistics using SPSS™. For the main analyses, we used multiple linear and logistic regression models to assess the unique and independent relations between the predictor variables (recruitment condition and exhibit version) and the outcome variables (engagement time, mathematical exhibit behaviors, and level of articulation of the mathematical relationships in the exhibit). The regression analyses allowed us to examine the relations between the recruitment conditions and the dependent variables, controlling for differences between the two versions of the exhibit (research question 1), and to explore potential interactions between recruitment condition and exhibit version (research question 2). Based on a review of residual plots, distance, leverage, influence measures, and collinearity diagnostics, there was no indication of outliers or violations of the assumptions of the regression models (Tabachnick & Fidell, 2007). All significance tests were two-tailed unless specified, using a critical value of 0.05. For each of the dependent variables, we first describe the regression model without the interaction between recruitment method and exhibit version included, and then describe the model with the interaction term.

Results

There were no statistically significant differences between the recruitment conditions based on average child age, total group size, proportion of adult females in group, proportion of child females in group, adult interviewee gender, proportion of adult interviewees who had previously visited the exhibition, or number of times adult interviewees reported having visited OMSI in the last six months.[1] There was a statistically significant difference between the two groups in the total number of children per family ($U = 1540.5$, $p = 0.007$). On average, families in the actively recruited group had more children (*Mdn* = 2, Range: 1–7) than families in the passively recruited group (*Mdn* = 1, Range: 1–4). Because visitor and group characteristics were

[1] Although visitors who had previously visited OMSI or the *Design Zone* exhibition might be expected to behave differently at the exhibit compared to other visitors, these participants were included in the analyses in order to support generalizability of study results to a broader range of museum visitors. There were no statistically significant differences in prior visits to OMSI or the exhibition across experimental conditions.

not significantly related to the outcome variables in the initial descriptive analyses, they were dropped from subsequent models.

Engagement Time

The mean engagement time for actively recruited visitors was 7.5 minutes (*SD* = 3.40, 95% *CI* = 6.67–8.36, *Mdn* = 6.85), compared to 2.8 minutes (*SD* = 2.01, 95% *CI* = 2.24–3.28, *Mdn* = 2.12) for passively recruited visitors. Engagement time was unintentionally not recorded for one group in the passively recruited condition and one group in the actively recruited condition. Therefore, the total sample size for subsequent analyses was 126.

To explore potential effects of recruitment method on visitor behaviors at the two exhibit versions, we used multiple linear regression analysis to evaluate the relations between recruitment condition, exhibit version, and visitor engagement time (Table 1). For recruitment, the passively recruited condition was coded as the reference category. In other words, results displayed in the table show the average difference in visitor engagement time of actively recruited visitors compared to passively recruited visitors, controlling for all other variables. For exhibit version, the reference category was the no computer condition, with the computer guide covered. In the first analysis (Model 1), recruitment condition and exhibit version combined explained approximately 48% of the variance in engagement time ($R^2 = 0.484$, $F(2,120) = 56.303$, $p < 0.001$). Controlling for differences between the computer and no computer condition, active recruitment was significantly and positively related to the length of time visitor groups spent at the exhibit ($B = 4.352$, $t(120) = 10.390$, $p < 0.001$). On average, actively recruited visitors spent approximately 4.4 minutes longer at the exhibit compared to passively recruited groups. As shown in Model 2, adding the interaction between exhibit version and cueing condition did not significantly increase the explanatory power of the model ($\Delta R^2 < 0.001$, $\Delta F =$

0.012, $p = 0.911$). In other words, the recruitment method appeared to be strongly related to visitor engagement time but did not influence comparisons of groups at the two versions of the exhibit.

Table 1. *Multiple linear regression predicting engagement time*

Variable	Model 1 (*B*)	Model 2 (*B*)
Constant	2.156***	2.132***
Active recruitment	4.352***	4.398***
Computer guide	1.150**	1.198*
Computer guide x active recruitment		-0.094
R^2	0.484	0.484
F	56.303***	37.230***
ΔR^2		0.000
ΔF		0.012

Note. Unstandardized regression coefficients shown. $n = 126$.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Total Laser Patterns

We also examined whether or not recruitment method was related to the nature of visitor interactions with the exhibit, as measured by the number of unique laser patterns created by visitor groups. The mean number of unique laser patterns in the actively recruited condition was 10.0 (*SD* = 3.53, 95% *CI* = 9.12–10.85). In comparison, the mean number of unique laser patterns for passively recruited visitor groups was 3.9 (*SD* = 3.24, 95% *CI* = 3.05–4.69).

As with engagement time, we conducted a multiple linear regression analysis with total laser patterns as the outcome variable (Table 2). Without the interaction term (Model 1), recruitment condition and exhibit version significantly predicted the total number of patterns created by visitor groups ($R^2 = 0.521$, $F(2,125) = 68.040$, $p < 0.001$). Both predictors were significantly and positively related to total laser patterns, jointly explaining approximately 52% of the variance in the outcome variable based on the model R^2. Controlling for differences between the computer and no computer conditions, actively recruiting visitor groups was associated with a 6.2 increase in the number of laser patterns created ($B = 6.189$, $t(125) = 10.989$, $p < 0.001$). Model 2 did not indicate a significant interaction between recruitment condition and exhibit version ($\Delta R^2 = 0.002$, $\Delta F = 0.044$, $p = 0.508$). In other words, the recruitment method appeared to be strongly related to the number of unique laser patterns created by visitors but did not influence comparisons of groups at the two versions of the exhibit. Results were similar when engagement time was included as an independent variable in the regression models.

Table 2. *Multiple linear regression predicting number of unique laser patterns*

Variable	Model 1 (*B*)	Model 2 (*B*)
Constant	2.634***	2.833***
Active recruitment	6.189***	5.814***
Computer guide	2.397***	2.010*
Computer guide x active recruitment		0.749
R^2	0.521	0.523
F	68.040***	45.303***
ΔR^2		0.002
ΔF		0.0440

Note. Unstandardized regression coefficients shown. $n = 128$.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Exhibit Messages

As a final set of analyses, we looked at how recruitment method potentially influenced visitor learning outcomes, operationalized as the level at which visitors articulated the mathematical relationships in the exhibit during the post-interaction interviews. Table 3 shows the percent of interview responses that were coded at each level, by recruitment condition and for all visitor groups. The median coding level was 2.5 (range: 1 – 5) for actively recruited groups and 2.0 (range: 1 – 5) for passively recruited groups. Across both conditions, approximately half of the responses were coded at level two (*identifying elements* category) and approximately half were coded at level three, four, or five (*relating two mathematical elements*, *relating three mathematical elements*, or *generalizing relationships*).

Table 3. *Percentage of interview response codes by recruitment condition*

	Recruitment condition		
Coding level	Active (n = 66)	Passive (n = 62)	Total (n = 128)
1. Not relevant	1.5	6.5	3.9
2. Identifying elements	48.5	51.6	50.0
3. Relating two elements	33.3	32.3	32.8
4. Relating three elements	13.6	8.1	10.9
5. Generalizing	3.0	1.6	2.3

We used a logistic regression to analyze the relations between level of articulation of exhibit messages, exhibit version, and recruitment condition (Table 4). For this analysis, a dichotomous dependent variable was created by distinguishing visitors who articulated the exhibit messages at a *high level* (levels 3 through 5) and those that articulated the messages at a *low level* (levels 1 through 2). Responses coded as levels 3 through 5 all discussed relationships between mathematical elements in the exhibit and so were assumed to be better aligned with the goals of the exhibition.

In contrast to previous analyses, the interaction term in Model 2 tended towards statistical significance (B = -1.436, Wald χ^2-statistic = 3.662, p = 0.056), indicating that the relation between exhibit version and articulation of mathematical relationships in the exhibit may be moderated by recruitment condition. Overall, including the interaction term, the predictors were significantly related to the log-odds of mathematical relationship articulation ($\chi^2(3) = 13.650$, $p = 0.003$, Cox & Snell $R^2 = 0.101$). In other words, recruitment method, experimental condition, and the interaction between the two were associated with the level at which visitors articulated the main messages of the exhibit after they interacted with the activity. Based on the Cox & Snell R^2, the predictors explained approximately 10% of the variance in the outcome variable.

Table 4. *Logistic regression predicting high-level articulation of mathematical relationships*

Variables	Model 1 (*B*)	Model 2 (*B*)
Constant	0.215 (1.240)	-0.134 (0.875)
Actively recruited	0.314 (1.369)	1.009 (2.743)†
Computer guide	-1.093 (0.335)**	-0.377 (0.686)
Computer guide x actively recruited		-1.436 (0.238)†
-2 log likelihood	166.737	163.014
Model χ^2	9.926**	13.650**
Cox & Snell R^2	0.075	0.101
Δ -2 log likelihood		3.723†

Note. Unstandardized logistic coefficients shown with odds ratios in parentheses. Reverse coding recruitment condition for Model 2, the regression coefficient for computer guide was: -1.814 (0.163), $p = 0.001$. $n = 128$.
†$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

As suggested by the interaction term, the relation between exhibit version and level at which interviewees articulated the mathematical relationships potentially differed between the two recruitment conditions. When visitors were not actively recruited (i.e., for the recruitment condition evaluated at zero in Model 2), the presence of the computer guide did not have a statistically significant impact on visitors' articulation of the mathematical relationships ($B = -0.377$, Wald χ^2-statistic = 0.533, $p = 0.466$). Reverse coding the recruitment variable (not shown in Table 4) indicated that the regression coefficient for the computer condition, evaluated for actively recruited visitor groups, was statistically significant (B = -1.814, Wald χ^2-statistic = 11.104, p = 0.001).[2] Results were similar when engagement time was included as an independent variable in the regression models.

To understand the implications of these results, we further explored the predicted values of the regression equation for the different experimental conditions. For passively recruited visitors, the predicted probabilities of being coded at a high level of articulation, based on the logistic regression equation, were 46.7% and 37.5% for the no computer and computer conditions, respectively. For the passively recruited group, the predicted probabilities of being coded in the high awareness category were 70.6% and 20.1% for the no computer and computer conditions, respectively. In other words, when visitors approached the exhibit on their own, there was no statistically significant difference in their articulation of the mathematical relationships in the exhibit based on whether or not they had access to the computer guide. When visitors were actively recruited, however, this difference was dramatic, with visitor groups that had access to the computer guide being much more likely to articulate the mathematical relationships at a high level compared to those who did not have access to the guide. (There were no statistically significant differences between the high and low articulation groups based on engagement time, $t(124) = 0.947$, p = 0.346, or total laser patterns, $t(126) = -0.140$, p = 0.889).

[2] In regression analysis, reverse coding categorical variables does not change the results of the analysis but rather is used to highlight or explore different aspects of the model. In Model 2 of Table 4, adding the interaction term means that the difference between the computer and no computer conditions is shown only for the passively recruited visitor group, since the coefficients represent changes when all other categories are evaluated at zero (i.e., the passively recruited visitor group, which represents the reference category for the recruitment condition in Model 2). Reverse coding the recruitment variable, therefore, reveals the experimental condition coefficient for the actively recruited group.

Discussion

In this study, we used data from a prior investigation (Pattison et al., 2012) to explore: (1) how, if at all, the behaviors and learning outcomes of museum visitors at an interactive exhibit change based on whether the visitors were actively recruited by a data collector or passively recruited through signage and (2) how, if at all, these approaches to recruitment influenced the impact of two different versions of the exhibit. The findings outlined above suggest that actively recruiting visitors increased the effect of participant reactivity and had a strong influence on visitor behaviors and learning outcomes. Comparing groups that had been actively recruited by a data collector with those that had been more passively recruited using posted signage, we found that actively recruited visitors spent significantly longer at the exhibit and engaged in significantly more mathematical exhibit behaviors, as indicated by the mean number of unique laser patterns created by visitor groups. There was also a marginally significant interaction between recruitment condition and exhibit version based on the level at which visitors articulated the mathematical relationships in the exhibit. Specifically, the difference between the two versions was statistically significant for actively recruited but not for passively recruited visitors. The associations between recruitment method and mathematical exhibit behaviors and articulation of mathematical relationships both remained significant even after controlling for engagement time, suggesting that participant reactivity influenced not only how long visitors spent at the exhibit but also the nature of their learning experiences. This research aligns with studies that have found significant participant reactivity effects (Eckmanns et al., 2006; Haidet et al., 2009; Spiers et al., 2000) and extends prior evidence (Serrell, 2000) that reactivity is a potential threat to the validity and generalizability of exhibit research.

One important question raised by the study is why additional contact with the data collector might lead to increased participant reactivity even if, as Gutwill's (2003) research suggests, the vast majority of visitor groups are aware that research is taking place when the appropriate signage is placed around an exhibit. Many researchers have argued that less obtrusive data collection methods can decrease reactivity (Haidet et al., 2009; Kazdin, 1979; Shadish et al., 2001; Spiers et al., 2000). Although more research is clearly needed to understand the social and psychological mechanisms underlying these findings, we speculate that the additional contact with the data collector makes this relationship more salient to visitors and increases pressure to behave in ways they believe are appropriate in the research context. Research in a variety of fields indicates that impression management and identity negotiation are fundamental aspects of human interaction (Falk, 2009; Gee, 2000; Goffman, 1981, 1982, 1990; Gumperz & Hymes, 1986; Norris, 2011; Penuel & Wertsch, 1995; Scollon, 1998) and scholars have argued that museums and science centers are important sites of identity enactment (Falk, 2009; Pattison & Dierking, 2012, 2013; Rounds, 2006). As in the study by Kohli and colleagues (2009), the more direct connection between visitors and the data collector may increase visitors' perceptions that they have something at stake and, therefore, increase participant reactivity effects, especially if data collectors are perceived as experts or if visitors feel their behaviors are being judged.

Because we did not collect detailed data on visitor interactions, we can only speculate on why recruitment condition potentially moderated the relation between the exhibit version and the level at which visitors articulated the mathematical relationships in the exhibit. As discussed by Pattison and colleagues (2012), the difference in the messages communicated by the computer guide and the explanatory label may partially explain this finding. The label contains a clear and concise summary of the mathematical relationship embedded in the activity: “the laser pattern depends on the relationship (ratio) between the speed of mirror one and the speed of mirror two.” The computer guide contains no such summary.

Actively recruited visitors may have felt more pressure to diligently attend to the most salient aspects of the exhibit. When the computer guide was available, this was likely the focus of adult visitor engagement, given the attracting power of technology in informal learning settings (Ucko & Ellenbogen, 2008). When the computer guide was not available, adult visitors may have been more likely to diligently attend to the explanatory label and, therefore, to benefit from the concise summary. This would explain the difference in the level at which visitors articulated the mathematical relationships within the actively recruited condition. On the other hand, when visitors were only passively recruited, they may have been less likely to attend diligently to any aspect of the exhibit, and thus the difference between the conditions with and without the computer guide would be less pronounced or nonexistent, as was seen in our data. This exploratory finding merits further research, since interactions between recruitment method and treatment conditions can potentially have important implications for study findings, as discussed below.

Implications for the Field

Although using a data collector to approach participants is a common and convenient method for recruiting visitors in both museum research and evaluation, the external validity of such studies, and the extent to which they represent naturalistic visitor behaviors, must be called into question. In this study, relative to the often brief and ephemeral nature of exhibit experiences, the influence of recruitment method on visitor behaviors was dramatic. Researchers and evaluators must consider these factors when designing studies and reporting findings.

Furthermore, for investigators conducting experimental work in these settings, it is not only important to consider the overall impact recruitment has on visitor learning behaviors but also the ways in which it might interact unexpectedly with other research variables.

Comparing findings across the two recruitment conditions highlights these complexities. If we had conducted the study using only actively recruited participants, we likely would have concluded that the design of the computer guide needed to be rethought, given its significant negative impact on visitors' articulation of the mathematical relationships. Furthermore, we also would have greatly overestimated the time visitors spent at the exhibit and the total number of patterns they created, potentially minimizing the perceived need for a scaffolding tool such as a computer guide.

However, looking at data from passively recruited participants, which arguably represent more naturalistic visitor behaviors, it appears that the computer guide had a significant positive impact on engagement time and number of mathematical behaviors and little to no negative impact on visitors' articulation of the mathematical relationships. These data suggest a much more positive interpretation of the computer guide, especially considering the short amount of time that passively recruited visitors spent at the exhibit without the guide and the relatively few number of laser patterns that they created.

Given these findings, museum researchers and evaluators should look for strategies to minimize the impact of reactivity, particularly when participants are likely to feel pressure to behave in particular ways or conform to social expectations.[3] Researchers have suggested a variety of approaches to minimizing participant reactivity, including avoiding pretests that provide hints about expected outcomes, conducting pilot observations to better understand

[3] In some cases, it may be desirable to use recruitment methods that motivate visitors to spend more time at and engage more deeply with exhibit activities, such as during formative evaluation when developers are interested in the impact of exhibit design elements under ideal conditions.

naturalistic behaviors, providing time for participants to acclimatize to the research context, using less obtrusive or threatening observation methods, and reducing experimenter interactions with participants (Haidet et al., 2009; Kazdin, 1979; Rosenthal & Rosnow, 1991; Shadish et al., 2001; Spiers et al., 2000). Research on surveys specifically suggests that self-administered questionnaires are less likely to produce participant reactivity effects compared to interviewer-administered methods (Groves et al., 2009). To date, the efficacy of these approaches has not been directly studied in museums.

Given the potential for participant reactivity to influence study findings, it is also important that researchers and evaluators clearly report their study methods, including specifics about how participants were recruited and how informed consent was conducted. If possible, investigators should directly assess potential participant reactivity effects, such as through a review of visitor behaviors when video data are available, post-data collection interviews or questionnaires with participants (Barnes, 2010; Haidet et al., 2009), comparisons with other research, or the direct, experimental assessment of recruitment effects, as in this study. Participant reactivity should be an important consideration as researchers and evaluators interpret results and make claims about the degree to which findings generalize to other contexts, including more naturalistic settings (Serrell, 2000).

Future Research

Studying recruitment in a free-choice environment is challenging and although the current study highlights the importance of such considerations, several limitations should be considered when interpreting results. First, in this study visitors could not be randomly assigned to passive or active recruitment since, by necessity, visitors in the passive recruitment condition self-selected to approach the exhibit. Although the two recruitment conditions were systematically rotated by time of day and day of week and all visitors in the study had chosen to enter the *Design Zone* exhibition, it is possible that some visitors in the actively recruited condition would not have interacted with the Laser Light Show exhibit, either by choice or because it was overlooked or bypassed during their visit. This leaves open the possibility that the two groups differed in some way not measured or controlled for in the study.

Second, other differences between recruitment conditions, beyond interactions with the researcher, may have contributed to observed differences between the two groups, including slight differences in language between the informed consent sign and the researcher recruitment script and differences in the timing of the incentive, which was offered to actively recruited participants before their interaction with the exhibit and to passively recruited visitors after their interaction but before the interview. Also, because we did not ask visitors in the passively recruited condition whether or not they were aware of the informed consent sign, it is not clear the extent to which complete lack of awareness of being observed explains the differences in behaviors and learning outcomes or rather more subtle differences in social pressure from different levels of contact with researchers.

Rainbolt and colleagues (2012) found similar challenges in their attempt to explore the impact of visitor recruitment on engagement time in a study of exhibit circulation mapping techniques. Ultimately, the results of their study were inconclusive, since recruitment method (unobtrusive observation without recruitment, and active recruitment) was confounded with treatment condition (visitors exploring the museum on their own and visitors being asked by researchers to map their visit as they toured the museum).

Given these challenges, more research is needed to determine the extent to which the findings from this study generalize to other exhibits, learning contexts, and outcome measures. In the meantime, the research highlights the potential importance of recruitment and sampling decisions in museum research and provides empirical guidance for how these decisions might influence study findings. Additional investigations are needed to understand the underlying mechanisms of participant reactivity and to explore the efficacy of strategies for minimizing its effects. More generally, museums and visitor studies researchers should continue to investigate the benefits and limitations of different methods and approaches as part of the broader agenda to strengthen the field's methodological toolkit for studying informal learning.

References

Allen, S., Gutwill, J. P., Perry, D., Garibay, C., Ellenbogen, K., Heimlich, J., … Klein, C. (2007). Research in museums: Coping with complexity. In J. H. Falk, L. D. Dierking, & S. Foutz (Eds.), *In principle, in practice: Museums as learning institutions* (pp. 44–56). Lanham, MD: AltaMira Press.

Barnes, B. R. (2010). The Hawthorne Effect in community trials in developing countries. *International Journal of Social Research Methodology*, *13*(4), 357–370. doi:10.1080/13645570903269096

Bernard, H. R. (2006). *Research methods in anthropology: Qualitative and quantitative approaches* (4th ed.). Lanham, MD: AltaMira Press.

Brewer, M. B., & Crano, W. D. (2013). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 11–26). New York, NY: Cambridge University Press.

Chiesa, M., & Hobbs, S. (2008). Making sense of social research: How useful is the Hawthorne Effect? *European Journal of Social Psychology*, *38*(1), 67–74. doi:10.1002/ejsp.401

Eckmanns, T., Bessert, J., Behnke, M., Gastmeier, P., & Ruden, H. (2006). Compliance with antiseptic hand rub use in intensive care units: The Hawthorne Effect. *Infection Control and Hospital Epidemiology*, *27*(9), 931–934. doi:10.1086/507294

Falk, J. H. (2009). *Identity and the museum visitor experience*. Walnut Creek, CA: Left Coast Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi:10.3758/BF03193146

Fernald, D. H., Coombs, L., DeAlleaume, L., West, D., & Parnes, B. (2012). An assessment of the Hawthorne Effect in practice-based research. *The Journal of the American Board of Family Medicine*, *25*(1), 83–86. doi:10.3122/jabfm.2012.01.110019

Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of Research in Education*, *25*(1), 99–125. doi:10.3102/0091732X025001099

Goffman, E. (1981). *Forms of talk*. Philadelphia, PA: University of Pennsylvania Press.

Goffman, E. (1982). *Interaction ritual: Essays on face-to-face behavior* (1st Pantheon Books ed.). New York, NY: Pantheon Books.

Goffman, E. (1990). *The presentation of self in everyday life*. New York, NY: Doubleday.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., & Singer, E. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: Wiley.

Gumperz, J. J., & Hymes, D. H. (Eds.). (1986). *Directions in sociolinguistics: The ethnography of communication*. Oxford, UK: Blackwell.

Gutwill, J. P. (2003). Gaining visitor consent for research II: Improving the posted-sign method. *Curator: The Museum Journal*, *46*(2), 228–235. doi:10.1111/j.2151-6952.2003.tb00088.x

Gutwill, J. P. (2006). Labels for open-ended exhibits: Using questions and suggestions to motivate physical activity. *Visitor Studies Today*, *9*(1), 1–9.

Haddad, N. F., Nation, J. R., & Williams, J. D. (1975). Programmed student achievement: A Hawthorne Effect? *Research in Higher Education*, *3*(4), 315–322. doi:10.1007/BF00991248

Haidet, K. K., Tate, J., Divirgilio-Thomas, D., Kolanowski, A., & Happ, M. B. (2009). Methods to improve reliability of video-recorded behavioral data. *Research in Nursing & Health*, *32*(4), 465–474. doi:10.1002/nur.20334

Jacob, T., Tennenbaum, D., Seilhamer, R. A., Bargiel, K., & Sharon, T. (1994). Reactivity effects during naturalistic observation of distressed and nondistressed families. *Journal of Family Psychology*, *8*(3), 354–363. doi:10.1037/0893-3200.8.3.354

Kazdin, A. E. (1979). Unobtrusive measures in behavioral assessment. *Journal of Applied Behavior Analysis*, *12*(4), 713–724. doi:10.1901/jaba.1979.12-713

Kohli, E., Ptak, J., Smith, R., Taylor, E., Talbot, E. A., & Kirkland, K. B. (2009). Variability in the Hawthorne Effect with regard to hand hygiene performance in high- and low-performing inpatient care units. *Infection Control and Hospital Epidemiology*, *30*(3), 222–225. doi:10.1086/595692

Landsberger, H. A. (1958). *Hawthorne revisited: Management and the worker, its critics, and developments in human relations in industry*. Ithaca, NY: Cornell University.

Maisto, S. A., Clifford, P. R., & Davis, C. M. (2007). Alcohol treatment research assessment exposure subject reactivity effects: Part II. Treatment engagement and involvement. *Journal of Studies on Alcohol and Drugs*, *68*(4), 529–533.

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. Washington, DC: National Academies Press.

Norris, S. (2011). *Identity in (inter)action: Introducing multimodal interaction analysis*. Berlin, Germany: De Gruyter Mouton.

Nurse Rainbolt, G., Benfield, J. A., & Loomis, R. J. (2012). Visitor self-report behavior mapping as a tool for recording exhibition circulation. *Visitor Studies*, *15*(2), 203–216. doi:10.1080/10645578.2012.715035

Pattison, S. A., & Dierking, L. D. (2012). Exploring staff facilitation that supports family learning. *Journal of Museum Education*, *37*(3), 69–80.

Pattison, S. A., & Dierking, L. D. (2013). Staff-mediated learning in museums: A social interaction perspective. *Visitor Studies*, *16*(2), 117–143. doi:10.1080/10645578.2013.767731

Pattison, S. A., Ewing, S., & Frey, A. K. (2012). Testing the impact of a computer guide on visitor learning behaviors at an interactive exhibit. *Visitor Studies*, *15*(2), 171–185. doi:10.1080/10645578.2012.715010

Penuel, W. R., & Wertsch, J. V. (1995). Vygotsky and identity formation: A sociocultural approach. *Educational Psychologist*, *30*(2), 83–92. doi:10.1207/s15326985ep3002_5

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York, NY: McGraw-Hill.

Rounds, J. (2006). Doing identity work in museums. *Curator: The Museum Journal*, *49*(2), 133–150. doi:10.1111/j.2151-6952.2006.tb00208.x

Scollon, R. (1998). *Mediated discourse as social interaction: A study of news discourse*. New York, NY: Longman.

Serrell, B. (2000). Does cueing visitors significantly increase the amount of time they spend in a museum exhibition? *Visitor Studies Today*, *3*(2), 3–6.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Spiers, J. A., Costantino, M., & Faucett, J. (2000). Video technology: Use in nursing research. *AAOHN Journal: Official Journal of the American Association of Occupational Health Nurses*, *48*(3), 119–124.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson/Allyn & Bacon.

Ucko, D. A., & Ellenbogen, K. M. (2008). Impact of technology on informal science learning. In D. W. Sunal, E. Wright, & C. Sundberg (Eds.), *The impact of the laboratory and technology on learning and teaching science K-16* (pp. 239–266). Charlotte, NC: IAP/Information Age Pub.

Weiss, B. H., O'Mahony, M., & Wichchukit, S. (2010). Various paired preference tests: Experimenter effect on "take home" choice. *Journal of Sensory Studies*, *25*(5), 778–790. doi:10.1111/j.1745-459X.2010.00309.x