

Interview on Evaluation in Informal Science Education: Sue Allen

Suggested Citation: Fu, A. (Interviewer) & Allen, S. (Interviewee). (2013, July 10). *Interview on evaluation in informal science education*. [Interview transcript]. Retrieved from <http://www.informalscienceevaluation.org/>

Interviewee: Sue Allen, Principal, Allen & Associates

Interviewer: Alice Fu, SK Partners

Note-taker: Lisa Peterson, SK Partners

Observer: Archana Kannan, SK Partners

Date and Time of Interview: July 10, 2013, 11:00am to 12:30pm (Pacific) (interview 1 of 2)

Location: Phone

As part of our efforts to understand current evaluation issues in informal science education (ISE), we conducted interviews with leaders in the field. We purposely selected a sample of individuals who could provide insights from a range of perspectives; collectively, they have experience with ISE and ISE evaluation as practitioners, evaluators, researchers, funders, and institutional leaders. Several participants generously agreed to share the transcripts from their interviews.

Please note:

- These are transcripts of oral interviews, *not* polished or written remarks prepared for publication.
- These transcripts have been edited for clarity, brevity, and ease of reading. Participants were also provided with the opportunity to remove any potentially sensitive material.
- The views or opinions expressed are solely of the individual interviewee and do not necessarily represent those of their affiliated organizations.
- We intend for these transcripts to serve primarily educational purposes. We believe that others may benefit (as we did) from the rich insights provided in these interviews.

Interviews were semi-structured: we used a protocol that ensured asking key questions in a comparable fashion across interviews, but there was ample flexibility to allow for interesting and unpredicted turns in conversation. The coverage and order of questions varied across interviews. Interview topics included but were not limited to participants' views on evaluation uses, methodologies, "best practices," and challenges. Interviews were conducted in-person or by phone, and each lasted approximately 90 minutes.

In these transcripts, the following conventions are used:

- Initials indicate who is speaking. **Blue text is used when interviewer is speaking.**
- *Italics indicate paraphrasing or researchers' comments/interpretations.*
- 'single quotes' indicate hypothesized thoughts or questions; e.g., And I asked 'what have you had done before? And what did you think of it? And what do you need?'
- - single dash indicates an interrupted thought or change in thought; e.g., It's just been - I was just so happy to have had that opportunity to work with them.
- ... ellipses indicate overlapping speech, deleted sections.
- [brackets] indicate non-verbal observations and other clarifications added by SK Partners.

A Note about this Interview:

This is the first of two interviews that SK Partners conducted with Sue Allen. Much of this first interview, and part of the second interview, focused on a summative evaluation that she conducted of the *Secrets of Circles* exhibition, developed by the Children's Discovery Museum of San Jose. For more information, see the evaluation report:

Allen, S. (2007). *Secrets of Circles summative evaluation report*. [Report prepared for Children's Discovery Museum of San Jose.] San Mateo, CA: Allen & Associates. Retrieved from http://informalscience.org/evaluation/ic-000-000-007-398/Secrets_of_Circles_Summative_Evaluation_Report

A key topic in our interviews was the design of the summative evaluation. After our interviews, Allen kindly provided the following summary:

I took a utilization-focused approach: working closely with stakeholders and trying to think of studies that would get to their questions, using benchmarks (from the literature or internal comparisons) wherever possible.

I started with the usual basic design of observing people's learning behaviors with a tracking & timing study, and then complementing that with interviews to get their reflections on their experience.

The evaluation design came quite pragmatically from the key questions the team had:

- Could they make the deep principles of embodied circles accessible to these young children? (which we assessed through their learning behaviors and their verbal reflections on the properties of circles)
- Could they reach a broad range of children (3-10)? (assessed through attraction, holding, and whether the main themes came through)
- Could they create something really inviting for Vietnamese and Latino visitors without having an exhibit that was about the Vietnamese or Latino cultures per se – it was fundamentally about circles. So could they make it culturally inviting through the design, and through the multilingual signage, and through the choice of some of the examples (like the round boat)? (assessed through interviews and any evidence of changes in behaviors)
- Was the multilingual signage readable, effective, embraced by all? (assessed through label use and through interviews)
- Could they get any evidence that design makes a difference? (assessed through interviews, though this only captured design features visitors were consciously aware of)

The interest in different age groups created one natural dimension for internal comparison. And the ethnicity of the visitors was another natural dimension for comparison. So I hoped that we'd get some interesting variations that would help us see overall learning impacts but also some interesting differences among visitor subgroups.

[BEGIN INTERVIEW]

AF: To warm up, can you spend a few minutes walking us through some personal highlights of your career, especially in relation to informal science education and evaluation?

SA: Yes. I was never trained formally in evaluation like so many of us. Kind of drifted into it. I had a background in science—in physics and astrophysics at college. I did my PhD in science education at Cal and then got a post doc to work at the Exploratorium.

It was the post doc that got me into informal science education. I was working part of my time there on the museum floor and part of my time with the teacher programs. And kind of read around and tried things, did a lot of experimentation with methods and approaches to assessing learning in this really challenging setting. And I'm still fascinated by that. It's a very hot issue. Just so intrigued by, 'how do you assess this?' and 'what really is learning?' and all those big, gnarly questions.

I spent 15 years at the Exploratorium and, in that time, established a department of visitor research and evaluation. We had a lot of big NSF grants come in, so there was a lot of work there. We had about 10 people on staff by the time I left, which is still going on very strong. I did that jointly with Kathleen McLean, I really want to give her credit. She was my boss and a great believer in research and evaluation and systematizing practitioner knowledge and testing it for evidence.

And then Jenni Martin approached me before I joined the project [evaluation of Secrets of Circles at Children's Discovery Museum of San Jose] and said, 'would I be interested in being the summative evaluator on it.' I had never done that before because, at the Exploratorium, the main model we had was that we did the formative evaluation in house and the summative we brought in independent consultants usually. That was a good split for us because, with the external people, we got to learn someone else's perspective, and usually, it was more fun for us to be in the trenches on the team. So that was the model we had, and the summatives were done outside. So when Jenni called, I thought, 'well I haven't done this before.' And in response to that request, I mean talking with her and knowing her a little and knowing Maureen Callanan and the people at UCSC, I thought 'wow, this is a really interesting project.' And so, in the end, I established my little consulting firm, Allen & Associates, to do this project [summative evaluation of Secrets of Circles].

AF: Oh interesting, I didn't realize that.

SA: So that was the reason. At that time, it was done in addition to my work at the Exploratorium. I was just so happy to have had that opportunity to work with them.

AF: And from there you went to NSF? Is that right?

SA: From the Exploratorium, yes.

AF: Can you say a few words about that experience?

SA: Yes, let's see, so I got a recruitment call from NSF, and I thought, 'this is an opportunity to go as a rotator without losing my job but just to see what the policy world was like, to be behind the scenes.' I also really liked the idea of being able to weigh in on what does get funded and some of the directions that the field was going. So I took a rotator position there and ended up terminating my employment at the Exploratorium. I was there for three years and, each year, I had a slightly different position. So I was a Program Director the first year, and then the second year I was the Cluster Coordinator, and the third year I was the Acting Division Director. What was great was I was always on a steep learning curve. What was tough was I felt like, by the time I got to the Division Director position, I was sort of out of contact with the work in the way that I really wanted to be, so that was one of the reasons I left at that point. It was: incredible insights, fantastic opportunity.

One of the things I would say about NSF is that I really appreciated its process. I feel very happy now to come back to the field and apply and feel that if I get declined or accepted, it's had many good eyes and much thoughtful process. Of course, NSF is staffed by mortals, and they are fallible and human and they have limited areas of expertise as we all do, but within that system, I just think there are a lovely set of checks and balances and processes. I understand now why it takes a year.

And I also learned about informal science, and I think it's an idea whose time has really come. I actually think it's a very exciting time for the field to be taken so seriously for the first time. I think a series of reports coming out, the president's interest, the PCAST [President's Council of Advisors on Science and Technology] report. A bunch of things really shifted. The role of choice has really become more fore-fronted in the political discourse, and so I think it's really not so much the infrastructure, but at least the issue and the potential role that informals can play. I guess there's still a lot of interest in that. So it was an exciting time to be there.

AF: That's wonderful. Thank you for that run through your history. It was helpful for contextualizing this project in particular. For Secrets of Circles, you said you were contacted by Jenni and it was different from what you had been doing at the Exploratorium because it was a summative project.

SA: Right. [SA laughs.] I'm now getting used to that kind of stretch as a consultant. You always feel like, 'I'm not sure I can do this project. Really? You want me to do this project'? One of the things was that it was a children's museum. At the Exploratorium, we do get people of all ages, but mostly we work with people we can interview comfortably. So that's age 10 and up, or sometimes maybe 7 and up, but very rarely below that. And the target age for this was 3 to 10, so these were younger children, and I didn't feel as familiar with that and with what parents and children do in those children's museums. It just seemed like a different setting.

Also, of course, there were even more issues around methods. One of the things was just stopping people. How do you stop people and get them to talk with you, when there are young kids in the group? [SA laughs.] It turned out to be very challenging. One of the things I did early on was to recognize that I could really reach out to colleagues who had expertise working with children's museums. So I contacted Lorrie Beaumont, and I contacted Cecilia Garibay. These were all people that I put in my acknowledgements. And also Veronica Garcia-Luis, who was already a close colleague at the Exploratorium. Cecilia and Veronica for their expertise in cross-cultural work and culturally responsive evaluation, and Lorrie for her work with children's museums in particular. So that was fabulous because I just went to them. I asked them if they would be willing to be sort of short-term consultants for me. They said, 'yes,' and they gave me wonderful wisdoms and were just very helpful. Plus, Mary Kidwell helped me design and pilot-test many of the instruments, and she had very good intuitions.

AF: It's funny you mention Cecilia. I have a question about that. One of the things we're trying to do as we talk to different people is understand where *Secrets of Circles* and its summative evaluation fit among all of the other things that were happening at CDM [Children's Discovery Museum of San Jose] at the time. From your perspective, how did your work relate to the work of Cecilia Garibay and her evaluation of the Vietnamese Audience Development Initiative, as well as the work of Maureen Callanan from UC Santa Cruz?

SA: That's interesting. Cecilia. Gosh, I don't remember being- you know, of course, all of this is my best memory.

AF: Yes, I know, it was several years ago. [AF laughs.]

SA: So caveat everything I'm saying. I don't remember talking with Cecilia much, if at all. That's interesting now, because you're saying her efforts. I'm not sure of the time frame on that. She would be the person who could tell you when she really did what with them because I don't remember her being deeply involved at the time.

AF: Yeah, it might have come later.

SA: Ok. In terms of Maureen, I can't remember exactly when I first ran into Maureen's work down there, but I'd sort of known for a while. We used to go down and occasionally give talks. And she would come up and give talks, and her research associates were part of our loose community at the Exploratorium. We just had a really nice sort of relationship with them in general for some years, I think.

And we knew, for this project, that Maureen was doing in-depth video analyses of the families' use of individual exhibits. They were setting up video cameras, doing the whole method of putting numbers on kids' backs, and all of that. They had developed these methods of doing this deep analysis in collaboration, I guess, with Jenni. I'm not sure about the history of that. I know Kevin Crowley worked with Maureen all those years ago. They had developed that sort of approach, and I think beyond that, they had started to put together a framework that Maureen shared with me. You know, I don't even have my old emails. I tried to see if I could find anything, but it's just too old.

I remember Maureen sending us, sort of notes from their reports as they were working with the team. I think they were doing research work and simultaneously providing formative evaluation as the exhibits were being developed. My feeling was that the formative was coming from Maureen and from in-house, and they were also doing the deep dives into what kinds of behaviors, what kinds of conversations were being observed real time. So, it was a good complement. And they had a particular framework too about how they looked at things that I think they had developed as a formative tool. They'd come up with some broad levels of analysis to help give fast turnaround. Talking to Maureen would be useful.

AF: Yes, we had a really wonderful conversation with her and she did talk a little bit about that. I think they call it 'blitz coding' -

SA: There you go.

AF: - that they do on the videos. So that's helpful, I just wanted to hear from your perspective.

SA: So I didn't participate in that, but I was aware of it. And I did talk with Maureen and with Jenni about 'what do you want the summative to be so that it is complementary, so that we're not duplicating.' So we didn't do anything with videos, we didn't do any deep dives. What we did was the tracking and timing piece and the interview pieces. And that made a good complementary set.

AF: Right. That's great. Actually my next question was going to be about main purpose or purposes of the summative evaluation as you saw it. We will talk more deeply about uses and

stakeholders later on in the interview, but if you could just help us frame it by taking a few sentences to talk about the main purpose, as you saw it, on this project.

SA: Obviously, they were mandated by NSF.

AF: Sure, yeah.

SA: My sense from talking with Jenni was that-, I asked her to send me the previous evaluations, because I'm very user-focused, like 'what's useful to you?' That's always where I come from, very utilization-focused approach to evaluation. And I asked, 'what have you had done before? And what did you think of it? And what do you need?' She sent me another evaluation study that had been done on one of their other exhibitions or areas. They hadn't done any, I think, tracking and timing before, and they hadn't done much in the way of interviewing. I think this might have been the first NSF project they'd had funded actually as a children's museum, so they were coming in without a whole lot of expectations. Very open to anything that seemed useful.

So they had these questions, and there was this realization that this combination of science being done at a children's museum was really under-studied. Children's museums were just not getting NSF funding much at that time. I think that's changed, and there's more people getting it now. But that was sort of an unusual combination. They had all these questions about, 'what is effective, what works for different ages.' And they were doing this creative combination of math, pure abstraction with this geometry idea of circles, with an engineering and circle function approach, and then with a cultural approach and their ongoing, obviously very deep, sincere mission to bring in Latino and Vietnamese audiences in particular because they were the largest ethnic groups in the San Jose area. So it was this real challenge, this project. Holy moly. How are you going to do all of this? And then to have it with such a wide range of ages, 3-10. Holy moly.

AF: Yes.

SA: That's a very broad set of ages. So, it was definitely a challenging project. And they were very open, but they had questions about their attempts to reach those audiences. The tri-lingual labels was a huge focus for them. And, were there differences in behaviors among the different age groups? How were the exhibits working? What evidence was there that these things were effective? And, of course, the cultural reference to the Vietnamese in particular. Was that going to be well-received? How would that fit with their larger mission to involve the community? And would that be appreciated and recognized? So it was just this really complex, multi-, multi-faceted thing.

AF: Yes, and that is a question we have, actually. As we read the report, we were really struck by the fact that there were all of these different elements, and you essentially conducted seven

separate studies that you reported on. So how do you decide what questions you wanted to pursue, and then how did you manage a project of that magnitude? [SA laughs and AF joins her.]

SA: I have no idea. [laughs] I had three fantastic advisors who helped me to get started when I was really freaked out at the beginning.

AF: These are people you mentioned earlier--Cecilia, Veronica, and Lorrie Beaumont.

SA: Yeah. I knew Lorrie Beaumont had done all this deep naturalistic work, but that wasn't what I really wanted to do, especially because Maureen was taking care of that deep dive. I wanted to get something across a large sample size, some kind of reflection to get people talking about circles, and not just watching their behaviors but reflecting on their experience. And then there was the thing about the two audiences, part of it just evolved.

I started out thinking the typical summative evaluation for exhibitions, as you know, is sort of an observational component, usually a tracking and timing study, and an interview component. That is the typical summative structure. I thought I could stick with that, given that we get different things from these. I didn't realize there were seven studies, and I wouldn't have set out to do seven studies. But I started out with that basic structure and then it became clear that we could capture, with the tracking and timing, all of their audiences of interest.

So we had a three-way split there in recruitment of young children, older children, and adults, so that we could do comparisons because I'm all about comparisons. I think if you have a number without context or a benchmark or anything, then what do you really know? So you could see over and over again in the studies where I was trying to say, 'what benchmarks do we have for comparison?' Or, 'how do the younger and the older compare, or how do they compare with the adults?' So at least internal comparisons, if you can't get an external benchmark.

AF: Yes.

SA: I think the main studies happened because we were hoping that we would get enough Vietnamese and Latino visitors that we would be able to just look at them as subgroups. But then the numbers were just too small.

AF: Sure.

SA: Then it became this ongoing, rolling, 'Ok, what do we do now? What do we do now?' - looking for ways to bring in these audiences. There were these special events.

With the Latino special events, I learned that the people we interviewed showed no particular differences, they seemed just like the other audiences that we'd had on weekends. That was interesting for me. One of the discussions that we had about these ethnic audiences was the usual one about the sensitivity around not wanting to say, 'Latinos do this versus Whites do this.' You don't want to stereotype and draw attention to the differences, but at the same time you are trying to identify those aspects which will appeal to those audiences that you are trying to recruit. So it's this delicate dance. I remember thinking when I saw that it was just the same data, 'Aw man, so we put all this effort into all these interviews and all these translations, and we're just getting the same thing again.' So, it was just sort of a methodological insight that if you don't have huge differences in behaviors, you end up not learning a whole lot.

As opposed to the Vietnamese study where there were so few Vietnamese [SA laughs] who came on the Vietnamese open day that we had to go to yet another plan and this special recruitment, which turned out to be fabulous. Actually, I think it might have been Cecilia who suggested that, or maybe it was Jenni. I don't remember. It wasn't me. I think my role was generally to say 'hey guys, we need to reach this audience, what opportunities do you have?' And they would say, 'oh there's this or oh there's that.' With the Vietnamese, we were so scraping the barrel to get people that we came up with this other idea, which was inviting this small group. We got what turned into this fabulous opportunity with this community leader. That turned out to be a very interesting structure. It's got pros and cons like any method. We ended up getting wonderfully rich data and a narrative about the data from her, but we only had 8 people and they were very particular kinds of families. So I'm always cautious about that study. There is a tradeoff I guess, depth vs. breadth. That was a strong selection effect. One just has to be careful because you read that, and it's so compelling that you think, 'oh, right, we need to watch out for Chinese hats.' But actually, we just have to remember, this was a very small subgroup of people.

AF: You mentioned depth vs. breadth. I guess you can think about that study in particular or the larger study as a whole, how do you weigh those trade-offs? Whether you want to dive in with a smaller set of studies or a smaller set of people vs. a broader look at things?

SA: Yeah, well, I always try to get both. The classic mixed-methods approach: you have something deep that gives you the texture of the learning experience and then you have something larger so you have an idea of how representative that is. That's always what I try to do. Again, because we knew that Maureen was handling the deep dive for the regular audience, I did more of the large numbers and sort of the median cut through it.

AF: Right. Another thing I wanted to follow up on, you talked about the importance of comparisons and the different kinds of comparisons that you highlighted in the report. One of the things that we noted was at one point you talked about considering pre-post comparisons, but it

was “beyond the scope of the evaluation.” Can you tell me a little bit more about that? That process of thinking through that alternative?

SA: One of the pieces of this was how to get at whether people were learning about circles, both the adults and the kids. What’s in the report is: yes, they could all talk about circles, they could make connections to their lives, they could recognize that the exhibition was about circles. But there is this part of me that is very cognizant that maybe they could tell you about circles before they came in.

AF: Mm hmm.

SA: What I wished I’d had was some kind of comparison group to say, ‘can you talk about circles before you see it?’ And do that comparison. There are ways to do that by getting people just before they go in vs. just after and so on. [Pause] But yeah, we ran out of time. There were only so many opportunities, and you have to get a certain sample size. We already had this three-way split of our data, because it was a stratified sample.

AF: Different ages.

SA: Right. I remember feeling like holy moly, there was a very, very strong time limit. The exhibition is just not going to be available longer than that. I think if it hadn’t gone touring, I would probably have been drawn irresistibly to some more studies, so I’m grateful in that respect that it was done. [AF laughs] It comes back to that benchmark thing. It’s great that people are doing these things and seeing these things, but compared to what?

AF: Yeah, can you maybe talk a little more about that. About weaving in those benchmarks and the importance of that to you?

SA: I think, as a field, we’re constantly up against this thing, ‘yeah, but what did they learn?’ It’s just a very deep, deep thing we’re up against. Part of our response is to say, ‘well, learning isn’t just reciting three mammals,’ or ‘learning isn’t just stating the five properties of circles.’ So we do that and I think that’s really true and really important that learning is all of these aspects of learning that the NRC document talks about. It’s affective, and it’s engagement, and it’s participation, and it’s discourse, and it’s activity, and it’s identity formation, and all of that. And much of it makes sense to assess in terms of a situated process.

But I’m a real believer in doing pre-post when you can. I like to have some process in my definition of learning and then some outcome that’s quantifiable with benchmarks. That’s why I think Beverly Serrell’s work with tracking and timing was so powerful – she gave us

benchmarks across a broad range of exhibits and exhibit types. So I was constantly trying to get at, ‘can we get a sense of evidence that something has been learned that wasn’t known before?’

For the tracking and timing, jumping back to that, one of the things I wished the field had was a mathematical model of the null hypothesis to compare your data to. So that would be a model of what visitors would theoretically do if you had exhibits with no real differences in their design. Like they would come in and see something by the front door, and then see the next thing and they’d go to that. It’d be like Brownian motion, boing, boing, boing, boing, mostly nearest neighbor, toward the door. With Secrets of Circles, I was quite disappointed that we had this really precise, really detailed tracking system, and then when we looked at the data, we saw people moving out across the room, mostly to the next exhibit, spreading out a bit and then coming back again to leave by the same door they came in.

I should say that Joyce Ma, another amazing person, designed the tracking software on a tablet way before anyone else was doing that. Coming back to your management question, the other piece that was so great was that Joyce is a brilliant tool designer at the Exploratorium, and she wanted us to be trying this because she was developing it. And Adam Klinger, who was already part of the staff at the Exploratorium, he was my right hand person, he did all the data collection for tracking and timing, and he is also a brilliant technical person, a great designer, very thoughtful, very detail-oriented, a super guy. So I had a super tool and super guy. [SA laughs] Our attention to detail was really almost obsessive compulsive. We probably had more precision in our data than any other study. I can’t imagine a more precise data set. But we didn’t know what to do with it to really use that precision.

So Joyce did this lovely visual representation where you could see how people were moving and we have four or so charts in the report so you can see those graphs. But what we need is compared to the null hypothesis because looking at those charts, I think what we had was the null. We had people just doing the bounce, bounce, bounce to the next exhibit in their field of view, and then back again toward the door.

AF: Interesting, and so that’s why, in an attempt to kind of find something to compare it to, that’s where you brought in Serrell’s measures?

SA: Right. So Serrell allows me to say, ‘Yes, this was very well used. Yes, this was very engaging.’ Certainly we knew that parents and children stayed together. And yes, parents interacted as well as children, which doesn’t always happen, so we knew we had something interesting. We could say those things, from the tracking and timing.

AF: This is so wonderful. It’s fun getting a chance to ask you all the questions that popped up in our heads as we were reading the report. You talked a little bit about how the study evolved as

you went along and there were various questions that the CDM staff wanted to answer. So I guess my question is about picking and choosing and prioritizing, and how do you decide which of those questions get your attention? Or do they all?

SA: [pauses] This is hard to remember because it's going way back and I don't have the documentation. I don't think they had a thousand questions. They had maybe 5 main areas of concern. The labels came up over and over again in the conversation. 'What about these tri-lingual labels, will they read them, will they appreciate them, should we be doing this?' And this whole circles thing. 'Could we connect cultural with science and with math? And could we connect the abstract with the engineering piece? Was that going to be effective?' And the age difference. And then, 'how will this work for the Vietnamese audience. Will they get it?' I mean there were sort of key things they would ask that you could just see were the big issues because Jenni would say them, and then when I would go to the meetings, other people would say them too.

I didn't do an evaluation matrix in a formal way, but I think I just laid out, 'here are the main blocks. We'll do tracking and timing and that's where you'll get this and this and this.' I always like to, as much as possible, sort of read the executive summary of the NSF report before we've even done anything to say, 'ok, this is what it's going to say, it's going to do say, '27% of the 3-5 year olds did this, then 80% of the 6-10s did that,' is that what you want?' I try to make it quite concrete, so people can say, 'ok, that would be great.' Part of that is driven by the lack of benchmarks. Because in that process, it helps them to think about what is good enough for them. What are they really going to be able to infer from all this data? Because I tend to be a data hoarder. I keep adding questions and wanting more data, so I can get out of control that way. So with the least encouragement from a client, it can get crazy. So, it is helpful to say, 'Ok, so here's what we're going to get from tracking and timing, and from interviews we will get some self reflection, and we'll get a sense of whether they noticed the design.'

There was this whole question about their response to the design. That was another big one; another of their big 5 was the response to the design. I'm thinking holy moly, how will you get at that? Design can be having a huge impact but it can be totally transparent to the consciousness of the visitor. So, how to get at that? That's sort of the fun of the problem solving. They want to know these deep things about the impact on the audience, and they want feedback on their design process and their design choices. How to give them that? So sitting down and thinking, 'Ok, how many ways could I get that? I guess I could ask people if they noticed. And use a general funnel structure.' I try to go with funnel structures: start out, 'did you notice anything about the design?' and then, 'did you notice anything that was Vietnamese-specific? Did you notice the bamboo?'

AF: All of this planning that's happening- At what point did you come on board? Was this planning happening along with the initial proposal to NSF?

SA: No, no, I came in after the grant.

AF: Ok.

SA: They were already pretty far along with their prototyping too, so they could show me things and give me a sense of what they had in mind. I would go down periodically. Not that often. They also had a change in staff. Peggy came on board instead of Michael, I think.

AF: Yes, we heard about that.

SA: She had her own ideas. I mean they're both amazing developers, but there was sort of a shift at that point. I really wanted to get in Peggy's head and I really wanted to get in Jenni's head. They were very articulate about what they wanted to know. So, for me, it had these steps of 'first tell me everything and let me understand what you are trying to do and what you want to know.' And then I go away and think again and then come back and say, 'here are pieces we can do, which of these feel most important to you? This kind of study or that kind of study would get at this or that.' They would say 'this sounds more important than that.' And then piloting each one of them and saying, 'ok, let's try this, this seems to work.' And then coming back.

They were also a fabulous client, because they were so excited by everything. They also gave me feedback, like 'this is fascinating, we'd really like to know more.' I would come down and give reports on what I was doing and how it was going. They were just so clearly eating up everything. I know Peggy read it all from front to back and I suspect Jenni did too. They really wanted to see my graphs, they asked questions. I felt like they were really going to suck up every piece of juice in the project and the study.

So there was the piloting, and then there was the whole getting personnel, which was another big piece, which you're probably going to ask me about, but I'll just tell you anyway. The piece about getting the interviewers. I think we all knew that there was no way I could successfully talk to a Vietnamese audience or even a Latino audience, I don't speak Spanish. Knowing from working with Cecilia and Veronica, that's absolutely key, to get someone who is a cultural insider. So I asked very early on, did they have anyone who was a native Vietnamese speaker and also Latino bilingual, and they came up with these three amazing people. I can't remember the details of that process, but Let Vu was the only person who was even remotely appropriate and she turned out to be just delightful and fabulous. And there were these two students who were both willing to come in and do the bilingual surveys with the regular audience as well as the Latino audience, and they did the translations. They translated the questions, and then they translated the data back to English so that I could still analyze it.

AF: Wow.

SA: Yeah. Which I know has its limitations. Ideally, you do the analysis in the native language too, but I couldn't do that, so that was the cost of working with me.

AF: That's helpful. We did have a question about how you pulled all those people together so thank you for that.

SA: I think I interviewed them and made sure I built a relationship with the interviewers. Made sure that they were excited by the project and that they seemed open, and they understood what this was going to entail. That they were willing to commit, that they were able to interview without leading people too much. I spent quite a bit of time working with them and explaining. And I trailed them and gave them feedback, trying to be a coach, so that they felt that I was really there supporting them, and as things came up they could ask me. So, for the first while, I was there to kind of get them up to speed. Then they were all so good I could just let them go and say, 'I'm on my phone, call me if things come up.'

AF: Right. That's great. Thank you for all these insights into how this study evolved and about how you pulled the team together. Thinking about how the study was designed and conducted, what were two of the biggest challenges that you faced?

SA: Definitely recruitment of these targeted groups that were not coming. Anytime you are trying to reach out to a group you don't know well yet, how slow that process is. I was just so in awe of the gentle persistence and enthusiasm that the CDM staff had for that. I just thought, wow, because over and over we had these challenges with recruitment and challenges with reaching the people and all the reasons that the Vietnamese and Latino audiences would say 'well we don't come because we don't know about it, we can't park, expensive,' and on and on and on. They [CDM] just kept going and saying 'no, we're going to do the outreach, be on the radio, we're going to make this happen.' So kudos to them, my gosh, they were so sincere and committed. But for me as an evaluator, it was very challenging to get enough people in the room to feel like I could really say anything about that community. As you saw, we ended up getting that community leader, Quyen Dinh, who served as an interpreter based on her experience. Fortunately, she was in education too. She was able to really give some fascinating insights on her experience, integrated over all this time, working with Vietnamese families, and how Vietnamese typically wouldn't show excessive pride in their language, and how they might not realize play is a part of learning. So I would say the cultural barrier of reaching that audience and feeling like it was done with enough people and with validity was very hard.

AF: Mm hmm

SA: And then working with young children. It was sort of like yet another cultural audience that I hadn't worked with. Almost the same thing. 'How do I reach these people? How do I handshake with them such that I am asking meaningful questions and getting meaningful responses? How to get children to reflect on what their experience was?' One of the things we used was this picture with all the photos of the exhibits on it. That was a powerful tool. It was also very powerful for multiple reasons that we were in an enclosed space. I think it made the experience more powerful and coherent for the visitors. I think we actually got better results. I think it was also less distracting. But also it was good because it allowed me to talk to the kids and the adults outside the space. So, at that point, they're not distracted and you also know they are going from recall, they're not looking at the thing and figuring it out as they're going.

AF: Oh interesting, uh huh.

SA: Once they're out of the space, then you have the recall problem. Right?

AF: Yeah.

SA: So I think this was a good way, to have this set of pictures and the children could say, 'oh yeah, I remember that one, I used that one, I used that one.' And then I could say, 'can you find the circle in that one that you liked?' So it ended up being, not quite an embedded assessment, but it wasn't horrific.

AF: Mm hmm [AF laughs]

SA: We did get a benchmark that 2/3 of them could do it, or 1/3, I forget how many it was, but they could recognize and talk a little bit about the ones they had seen. They could see if they could find a circle, and then they could talk about circles in their lives. So, I think it was as good as it could have been in terms of getting their reflections. So I was proud of that and it was very tough.

AF: It's like you're predicting my questions. My very next question on my paper here is about this challenge of rigorously meeting your evaluation needs while remaining authentic to the free-choice nature of the informal learning experience. So I was going to ask about your thoughts on this issue and how it played out in *Secrets of Circles*. Do you have anything you would like to add to what you just said?

SA: Obviously, the tracking and timing is completely invisible, so that's one of the lovely things about it. I think that's why it has this long history and is so compelling as a method. Nobody ever knew we were tracking them. Actually, that's not true because then we stopped them at the end and asked them some questions as they were leaving, but minimally invasive.

The interviews, I guess, you come to your own level of comfort around how uncomfortable you are willing to make people be. I know some evaluators who will just stop people and say, 'you need to do this, you will take this interview now.' [AF, AK, & SA laugh.] And then there are other people who are just the other way, 'if you'd like to stop that would be great.' So, a lot of these decisions, how hard you probe, how many times you come back to something, you're just never going to get everybody giving an equal answer to every question. It's just a basic social science question I guess, 'how open-ended should your semi-structured interview be?' I'm trying to be in a balance I think. In terms of the write-up, just articulating where I was in that balance, giving a lot of detail so that people can say, 'ah, yes, this was a finding, but bear in mind that a bunch of people didn't even respond to that question,' or whatever.

For example, just to stop to interview the kids, I had to have something to occupy every child. Sometimes you can split up the family and the father and one child will go one way, and the mother and the other child will go the other way, but you have to be ready to have a plan for keeping that entire family engaged for 10 minutes or 15 minutes while you talk to the child. That's when I started coming up with all these ways of having things for the other kids to do. Or interviewing all the kids, even ones outside my target age range, so that they all felt like they were participating. But then keeping the parents clear that their time will come, that this is not for them right now. All those ways you try to guide the conversation to feel comfortable, to be clear about people's roles and to be quick. That's the other thing, you've got to be quick. They don't have a lot of time. So you come up with these sort of creative ways, you offer them bribes [SA and AF laugh], you give them toys to play with at the time, you try to be as clear as possible.

Some questions turned out to be duds. And some along the way you add in. Like somebody suggested, 'why don't you ask the parent what the child is learning?' What a genius idea. Because of course parents are articulate about that. It's a combination of articulateness and comfort. What can you get at? That's the low hanging fruit. We added that in halfway through and, sure enough, got some great data on that. I wish I had that at the beginning. But it's always this tradeoff. You're asking a lot of questions so the piloting is always the absolutely essential part of the whole thing.

AF: Right.

SA: You put out a lot of things and you try it and if questions don't seem to be helping you, drop them. Or if they are repetitive. You just recognize there is only so much time you can spend with someone and what is most important to get.

AF: Yes. [AF laughs] You are talking about our interview now. We have so many questions we want to ask of you, so we have had to make those hard choices too. What is absolutely essential with a limited amount of time?

SA: Yeah. It's also kind of nice for me because I can easily get into perfectionistic drama about all of this. It's really good for me to be reminded and to come back. Actually, the goals of the team helped to ground me too. 'Let that go. You're not going to get that, just let it go.'

AF: Interesting. OK.

SA: What is really powerful? What is really important? Really, I mean I felt like I could have gone on writing more and more. I had to make decisions to keep it down to 150 pages or whatever it was. There is so much going on and so much complexity and so much intriguing data that it's like I need someone to reach out to and say, 'ok, remind me here what is enough, and what's core and what is going to be useful to you?'

AF: Right. OK. You mentioned you came in after they started prototyping; they had gotten the grant from NSF. How did you prepare yourself for this evaluation? You already mentioned some things.

SA: Mostly through human contacts. Certainly setting aside the time to figure out the timeframe was a big deal. I wanted to be really clear on the scope and the timing. I'm very relationship focused in the sense that I wanted Jenni to feel that I was attentive to her, and that we had an agreement that was clear from both sides and that wasn't going to go south half way through. Because it was my first gig you know. And I was clear about that with her, 'I haven't done this before, so let's sort of figure it out as we go.' There was all the financial side of things, like how were they going to pay me? How was I going to invoice them? How were we going to feel like we had something that was flexible enough to be responsive to evolving changes as we needed, but also stable enough that I could plan around that and feel like I wasn't going to be doing a ton of work that ended up not getting covered because of my time. And also feeling like I could get the staff I needed in place. Once I got Adam to do the data collection and had the 3 interviewers or at least a sense that they were possible, then I felt much more secure that I wasn't going to be the one on the ground doing it all. And I looked over some literature on children's museums and interviewing techniques. But most of it was contacting people, because again you don't have a whole lot of time. So I went to the people that I knew and trusted to help with that. That was my consultant solution.

AF: Yup.

SA: It was about going down there, meeting with Peggy, meeting with people. I try to be incredibly respectful of staff at institutions I work at, from front line to director, everybody. And very appreciative of what they do. The staff there ended up doing me all kinds of favors.

AF: Sure, yeah.

SA: When the bubble solution didn't show up. We had some small catastrophes along the way. They just stepped in in this wonderful way because they knew me and we had a trusting relationship. So I think I tend to invest in building relationships as much as possible. And it was easy with them, they were great. I don't know, what other aspects of preparation are there?

AF: That was really helpful. You talked about building relationships and investing in the time to talk to people. One aspect of preparing for the evaluation is learning about the Secrets of Circles project itself. How did you-

SA: Oh yeah. I read the original proposal, and I would ask them all about the changes. I would highlight everything, go through things with a fine-toothed comb, and then I would see every prototype. I wanted to feel that I really understood the big picture of the project as well as the details and even the evolution, like what were they were figuring out along the way? But I relied a lot on their self report.

AF: On the team's, from CDM?

SA: Yeah, yeah. The scale of it was good because, basically, there was Jenni and Peggy and Tina, who was the designer. But Tina and Peggy were very close, so Peggy knew what was going on for Tina. Peggy was just fascinated by everything and wanted to know a lot of things, so I ended up talking with Peggy a lot. And Jenni. And Tina. But I didn't have to talk to 17 people.

AF: Right.

SA: It was a few key stakeholders. I got the design piece, and I got the exhibit piece, and I got the mission piece. So I feel like I got different pieces. Also, it was a coherent team, which was such a blessing, because some of the projects out there people are pulling in different directions. That makes it very challenging because you're never quite sure. It was a tight, coherent, strong team, which really helped.

AF: Thinking about Secrets of Circles and its origins and underlying rationale, did it have, when you were doing your digging, did it have an explicit theory of action or logic model or something similar to that?

SA: Not that I recall. I don't think they had a logic model. No, actually, I don't think they did. I don't know what their early thinking was. I think it rested to a large degree on the great intuitions of their exhibit developers and their strong programmatic backup for the cultural. They had done these cultural events with Vietnamese and Latino audiences before. This wasn't their first time. They had this trajectory that they were on. This was just part of it from that perspective. I don't think I ever had an explicit logic model. There's another whole discussion we could have, logic models.

AF: Yes. [laughs]

SA: I'm not a great fan of logic models. Maybe I don't use them well enough, but I feel like logic models can often just be sort of a laundry list, sort of an accounting of 'here's what we put in the pot and here's what came out at the end.' I think the interesting stuff is sort of the causal reasoning or linkages in the model, and those are often not explicit. So I don't know, I certainly didn't push them to do a logic model.

AF: I guess, then, I'm wondering if you didn't push them to do a logic model and you did all this homework and talked to the team, I'm wondering about how you construct the logic behind the intervention for yourself? Or if you thought about what those causal linkages might be that you just talked about?

SA: Yeah. That's an interesting question. [pause] [sighs] I don't think I did and maybe this is the part where I feel like it wasn't the greatest report and it wasn't the greatest study. It did some ground-up work, but I don't feel like it had a strong evaluation theory behind it. It was a pragmatic study of like, 'these are the questions, and how can I best get at those? And what does it seem to show?' I didn't build it into some larger construct of 'this is the major finding from this.' That is a weakness I think that could have been stronger.

AF: Yeah, that's helpful. That is something we've struggled with and how much of that is useful to do. What is appropriate given different situations? I guess we're just struggling too with where, what the role of that kind of work is in evaluation, and all the different kinds of evaluations that occur.

SA: Right. For me, part of it is also cultural, working with CDM, especially early on with Michael. Coming from the Exploratorium, there is quite a similar culture there. Michael Oppenheimer is the son of Frank. I was very cautious, which is something I've learned at the Explo, about forcing something into a certain form or being seen as a bean counter or insisting on my own framework. That does not go down well at the Explo, so that was sort of my natural

approach, is to have as light a touch as I could with things that could be seen as evaluation baggage.

AF: Yup.

SA: There were enough mid-size nuggets in there that it was interesting for me. I get the major half dozen goals and I think there are interesting questions in those. And I like the design focus because it's like design-based research, having some sense of their principles. I would push Peggy, as a developer, to say, 'what do you mean by 'secret of circles?'' Because it's such an interesting idea, and it wasn't clear to me what she meant. She would say 'it's what makes a circle able to do what it does. Things like symmetry and smoothness and roundness.' Then I got it. Ok, so that's a very deep goal from a STEM perspective: to connect the abstraction of a geometrical shape into its functional properties in an engineering context. Like, 'ah, I got it!' And then I could see how far did they go in that area. Well, they went part-way. They got people to see the circles, they got people to make connections to the engineering contexts in their daily lives, but they didn't get people to make that link as much as Peggy could, to say 'it's the symmetry properties that allow it to do x.' The people didn't do that. Visitors didn't make that deep connection, which I'm not surprised. So that was certainly a piece that was compelling.

AF: That's very helpful. Thank you. I'm going to switch gears a little bit and ask you to respond to the following: one common definition of evaluation is "the systematic investigation of worth or merit of an object." So I'm wondering, what does the word "worth" mean to you?

SA: There is worth to the field, there is worth to the various stakeholders. You're determining the worth to the visitor audience, to the institutional audience, and to the field. This is an NSF funded project and it advances the field. That puts another layer on it, which is 'what are some principles, or semi-principles or proto-principles, that you could say you've learned?' That's the other distinction I think. Something having worth as a direct service product vs. as a learning product. I think this was an ideal project which had both.

Because I think it was actually very successful as an investment in direct service. I think it reached a lot of people in a successful way, it traveled successfully, it had a lot of spin offs programmatically, it reached a lot of people, and so on. It certainly was well received by those Vietnamese and Latino audiences that I talked to. So the quality of the reception by the audience, and the deeper impacts, like people did seem to learn some science from it.

And then in terms of the principles and for the broader field, the professionals who are doing similar things, hopefully there are themes in that report, surprising things. For example, I think it was really surprising that the vast majority of the visitors, even the English-only speakers, supported this multi-lingual approach. I thought that was a surprising and really important

finding that I would want people to know about, that I've quoted since then. Almost all of them felt that. At the same time, it raises some interesting things about how you do it, like you keep these labels really short, and that allows you to do multi-lingual and take up a reasonable amount of space.

AF: Right.

SA: There were some very sweet qualities to their actual design solution. I think I rambled there.

AF: No, that was very helpful. I heard you say a number of things: worth to different audiences, to visitors, to the institution, and to the field. And because of this last audience of the field, pulling out some of what you called proto-principles that people could learn from.

SA: And benchmarks. I would also add benchmarks.

AF: And benchmarks. OK.

SA: The next person who comes along with something similar can look at the study and say, 'oh, ok, the engagement levels were this,' or 'the understanding levels were this.' I like to put out the actual data because that will allow other people to look at it for benchmark purposes later.

AF: You've done some of the connecting that I was going to ask you to do in this next question. From your perspective, what is your role, if any, in investigating what you defined as worth in the summative evaluation?

SA: I'm not sure exactly what you're asking but I would say we've had evaluators come to the Exploratorium who have actually said, 'You've got the data. I just collect the data and I synthesize it and I write it in a report.' And I have said, 'where is your sense of, how well did we do?' [SA laughs] They would say 'that's not my job.' I really don't agree with that. I really think it's a key part of what evaluators can bring, especially those who go across multiple projects and are familiar with a range of projects. That's one of the things they can contribute to institutions who only see their own work; evaluators can say 'this is typical, and this is not typical,' even if they're not writing up full-on peer reviewed journal articles. This is part of the grey literature problem of our field-

AF: Yes.

SA: is to know what typical benchmarks are for these things, really allows an evaluator to say 'well, you did pretty well.' If I hadn't had Beverly Serrell's study, then what would I have said? 'Well, we got this fraction, it looks pretty good to me, but I don't really know.' So I think having

benchmarks helps a lot to see how you're doing compared with the rest of the field. And that's also why I end up being such a proponent of shared measures across studies in general because then you can really say 'we measured the same thing.' I mean, it's not perfect, but at least it's closer. I think what Serrell contributed was the value of a shared instrument. So you can say, 'oh you found this, we found that, and why.' You can start having those discussions. If everything is individualized, you can't do that.

But anyway, I do think evaluators can use their own experiences and the grey literature to characterize the worth of the project, even if there isn't some external theoretical standard of value that you set at the outset. And I think it's important to name the weaknesses as well as the strengths. There's a sort of three-way collusion that happens between evaluators and organizations and funders, where it's in everyone's interest for the evaluator to say 'this was great.' So that's a systemic problem, because then we aren't really learning. Beverly Serrell did a skit one time at the VSA conference-

AF: [laughs] Oh really?

SA: that was a generic summative evaluation, and it said, 'lots of visitors came, many expressed interest but said they weren't experts, they understood the general theme mostly but they did not understand many of the details, and they had a generally enjoyable experience.' [laughter] So can we move beyond that? We need to be saying, 'here are the strengths, here are the failures, and here are the interesting issues.' I think evaluators can bring some of that.

AF: That's great. That is one of the things that we have been wrestling with. I think we all individually sit at different places on what the responsibility of an evaluator is, and we are trying to come together and find a place that we're all comfortable with.

SA: Well, I think as long as you are clear about when you're doing what. It's like the discussion in the results section of a paper. Here are the results, and then here is my discussion. And people can take issue with my interpretation and say, 'well Sue's benchmark is too high or too low.' But I've separated out 'here are my findings, and then here is my value judgment.' I just think we need to go there and name that that's what we're doing.

AF: Yeah. I'm realizing we only have a couple of minutes left, so I'm going to skip ahead to one of the last questions we had. A project like this covers a lot of ground, and you talked about things you would like to see in a report. I guess we are wondering, more broadly, how do you decide what to put in a report and what to leave out? And how to frame it in a way that is useful for its audience?

SA: [sighs] I think it's a great question. I'm not sure if I did a good job. I think one way is to give multiple perspectives, which I kind of did in that report. There is a summary and then there's the discussion at the end, and there are sort of little pieces on each, kind of cut by study. And then it's also bringing it back to the big questions in the recommendations section. Different pieces can allow, 'the answer to the funder's questions is in this piece, and the summaries of what I think is interesting in this piece.' So I think multiple perspectives is one way to do it.

AF: Mm hmm.

SA: In general, there are always the goals of the institution, of the client. And then there are always things that nobody expected to see, or things that grab you as an evaluator, you get fascinated by something. I think that has to be a dialectic. It has to be something where you have space for both.

AF: Yes. Ok. Thank you so, so much. I think I echo the thoughts of everybody in this room. This was really fascinating. We read through the report, and we had all of these questions about how these things get pulled off and how decisions are made along the way. So getting the chance to talk to you and ask those questions is a real treat.

SA: Thank you. It's actually super fun to talk about it. It's really a treat for me to revisit it with you. You guys are very kind and very supportive. [all laugh]

[END INTERVIEW]