

## Interview on Evaluation in Informal Science Education: Joshua Gutwill

**Suggested Citation:** Fu, A. (Interviewer) & Gutwill, J. (Interviewee). (2013, July 25). *Interview on evaluation in informal science education*. [Interview transcript]. Retrieved from <http://www.informalscienceevaluation.org/>

Interviewee: Joshua Gutwill, Director of Visitor Research and Evaluation, Exploratorium

Interviewer: Alice Fu, SK Partners

Note-taker: Lisa Peterson, SK Partners

Date and Time of Interview: July 25, 2013, 10:00am to 11:30am (Pacific) (interview 1 of 2)

Location: Exploratorium (in person)

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 Joshua Gutwill. Parts of both interviews focused on the Exploratorium's Group Inquiry by Visitors at Exhibits (GIVE) project. For more information, see the GIVE project website:

[http://www.exploratorium.edu/vre/visitor\\_research/give//index.html](http://www.exploratorium.edu/vre/visitor_research/give//index.html)

[BEGIN INTERVIEW]

AF: To warm up, kind of a broad question, can you spend just a few minutes walking us through some personal highlights of your career, especially in relation to informal science education and research and evaluation? How did you get to where you are?

JG: Sure, sure. I'll try. I'm trying to think, where do I start? I guess I'll start in college. I majored in physics; actually, I was a double major in physics and history. As I was into my junior year of physics, I was doing a lot of TAing, and two things were happening. One was that I started noticing a lot of the students I was TAing were having trouble in similar areas in understanding mechanics. At the same time, I was sort of feeling like the physics itself that I was doing was becoming more and more about math and less and less about being able to imagine really what was happening in the world, a lot of quantum stuff. So I was kind of becoming disenchanted with physics. Like, well, I don't think I want to go to graduate school in physics. I was getting interested in physics education essentially.

A professor named John Clement came and gave a talk. He's at U Mass, Amherst. He gave a talk to our physics department about misconceptions in physics. It was like, 'ahhhhhhh,' hear angels sing. This is so interesting, and it's just what I'm seeing in the students I'm TAing, you know, the freshmen. So I got interested in that, I went out to U Mass, I talked to him, and I basically got interested in cognitive science that way. I was fully thinking I would just go study with him, but he said 'you should look at some other programs too. I'll give you a full ride and whatever,' but he pointed me to Berkeley, and he was right. There were a lot of faculty working on this, and he was kind of the only one at U Mass at the time. So I looked at it, and I had never been outside of Massachusetts, so I thought, 'let me try it.' So I came out to Cal, and I did grad school at Cal in education in math, science and technology. I studied with Barbara White and Andy diSessa and John Frederiksen and George Lakoff. As I was finishing, it was the mid- 90s, and it was sort of a similar thing. Two-thirds of the way through, I realized 'I don't think I want to be a professor, I don't think I want to do academia.' I really thought I wanted to be an academic when I went in, but seeing academia up close, I thought 'I don't think I really want to do this.'

In the mid-90s, there was an explosion because of NSF requirements for evaluators of science education projects. So I started looking around at evaluation gigs. I ended up taking one as the director of assessment in the chemistry department at Cal, working for a consortium of universities that were trying to improve freshman chemistry - how do we move away from the didactic lecture model and into something more inquiry-based and group-based? And that was cool. I did that for four years, and I did a bunch of really fun studies, and found some cool stuff, published, and all that. I started to see, oooh, there's something quasi-academic here where I'm not actually an academic, but I can still publish, and I can still go to conferences and give talks. And I thought this is cool.

At the same time, Sue Allen, who was one of my closest friends and colleagues 'cuz she and I were both Barbara White's students at Cal at the same time, she came to the Exploratorium as a post doc. She won a MacArthur post doc and she came over here with that and basically started to build demand here for formative evaluation and even some thoughts about research. And so, as I was doing my thing, she was kind of building a department over here. As my thing was winding down, she posted a position for a senior researcher, a PhD level person. I applied, as did a couple of other colleagues of mine. Sue was really careful, she recused herself from all the decision-making [laughter from JG and AF]. She did not want any bias. And she also knew, rightly, whoever got the job, she didn't want them to feel like they got it because she knew them. She knew all of us, she knew all three candidates. It was a grueling interview process. It was really tough.

AF: Oh gosh.

JG: But one of the things in the interview process was to go out on the floor with her and someone else and look at a couple exhibits, and basically imagine yourself as a visitor and tell me all the things that are hard about using this exhibit. And it was sooo fun. I had never done anything like that - I had never been to the Exploratorium before.

AF: Oh wow.

JG: Yeah. I had really gotten interested in Don Norman's work on affordances and design and all that came out at the exhibits in the interview. I said things like 'but this thing is aimed up and it should be aimed to the left because you want this one aimed to the right'. I just had all these thoughts about how this design could be improved. And it was really, really fun.

And that was the beginning of my interest in physicality. I'd been reading about it but it was the first time - cuz' all my graduate work was all with computer models, simulations that students would use, and here was a physical object that people would interact with and learn through physical interaction. That sparked a real interest that I hadn't ever explored before. And I'm still hooked on that. Even outside of evaluation, I love thinking with the developers about how to design an exhibit. 'How do we want to lay this thing out? How do we want to set it up? Thinking about salience, 'What is the most salient thing when a person first walks up?' Then, how do you layer things so that they are less salient, but they are still there so that after I've got my first thing, I can dig deeper. I just have gotten really interested in that stuff.

So anyway, I was a Senior Researcher for 10 years. And then Sue left to go to the National Science Foundation. She became a program officer and then moved up the ranks there. When she left I became the Acting Director here. After 2 years she had to say whether she was coming back or not. She said 'I'm not coming back' and I became the Director of Visitor Research. So that's how I got here, in a nutshell.

AF: So when did you join the Exploratorium?

JG: 1998. Yeah. It's going to be 15 years in a month.

AF: That's amazing.

JG: And what's really weird is I'm considered a middling. I'm not even an old-timer.

AF: That's amazing. So there's people who have been here for decades?

JG: 30. 25-30 years, that's the old timers. And I'm like the mid-range. [AF laughs].

AF: Okay, so Director of Visitor Research and Evaluation. So, let's move to something that's rather fundamental. What do these terms mean to you? What does research mean? What does evaluation mean?

JG: So, for me, research and evaluation utilize a lot of the same methods but they have really different goals. The goal of evaluation is to learn something about 'this thing', whether it's an exhibit or a program or a curriculum or whatever. In a formative evaluation, the purpose is to learn something about "this thing" in order to improve it. 'Let's make it the best exhibit it can be.' In a summative evaluation, it is 'let's learn about this thing to see if we reached our goals. We set out to do x. Did we actually accomplish x, or to what extent did we accomplish x?' So for evaluation it is really all about "the thing." For research, it's really using "the thing" to learn about something more general, broader. In my work, it's to learn about learning. So I want to gain knowledge that I can share with others in the field about how people learn. Or a lot of the work that we do is a little more specific than that. It's what kinds of designs tend to foster what types of learning? So it's how design influences learning. Like I said, the methods can be really similar--you can do interviews in both, you can videotape people in both, and code the video for different behaviors or utterances, or interactions. You can survey people in both, you can focus group. Those methods are interchangeable; it's just *why you are doing it*.

And there's also kind of a continuum from direct applicability, something that practitioners would like [snap] instantly take up, is more in that evaluative realm. Something that is more general, but maybe less directly applicable, might be a little more theoretical, in the research realm. Something researchers would be more interested in than practitioners. But what we go for is the sweet spot, where we're talking about design and how it influences learning, so often everybody is interested. I mean, I don't know that everybody is interested, [JG and AF laugh] but hopefully it will appeal to both researchers and practitioners--like practitioners are interested in the designs because they want this kind of learning or that kind of learning to happen. When I say 'this kind, that kind' it is not so much about content, like physics or chemistry, but more the type of learning. Is it about building interest, is it about practicing skills, is it about conceptual understanding- that kind of stuff. Anyway, practitioners are interested in those designs and how they might come to affect learning. Researchers might be more interested in the kinds of learning we're looking at and how that fits with other models of learning.

AF: Do you have projects that will sometimes do both? For both evaluative and research purposes?

JG: Yeah, so there is a blend. The Going APE project had a lot of this. For example, we were looking at designs that would really foster self-directed learning. And the GIVE project also had it: where APE was designs of exhibits, GIVE was designs of these little programs or little games that you can play at an exhibit, how the design of those games affects learning. There are certainly practitioners who want to know all about Juicy Question and are interested in how Juicy Question is different from Hands Off because Hands Off didn't work as well as Juicy Question. We were very careful about how they were different from each other. They were very similar in this, this, this and this way, but they differed in this one way. But then we were looking at 'how does it affect inquiry?' People who study inquiry might be interested in 'what kinds of measures did you use for inquiry? How did you define inquiry? What does it look like?' Same goes for, in the APE project, 'What is self-directed learning? How did you measure that? What does it look like in this context?' Trying to have both pieces--the design and the learning outcomes.

AF: Can you talk a little bit about internal vs. external and what you do internally vs. externally? I think my understanding is that you do formative evaluation and lot of the research projects internally, and sometimes you have to do summative evaluations on perhaps those projects or other projects and you go externally.

JG: So who does what?

AF: Yeah. Exactly.

JG: So internally we do a ton of formative evaluation, some front-end evaluation, and we do research. And then externally, we always do summative evaluation. All of our summative evaluations are with external evaluators. It's not exactly mandated by NSF or by the funders, but it's very strongly encouraged. They have this belief that external evaluators are more objective. And I do think it's true. They are further from the design process so they're not as aware of all the thinking that went into the design, which can be actually be a benefit. They don't automatically fill in 'oh I see what this visitor must be doing because we were trying to get them to do that'. They can say 'what's this visitor doing?' and really ask the question.

But I have always bristled a little bit at the notion that internal folks are less objective in a sort of biased way like 'I want this to succeed.' Only because I think you can make an argument that, there is definite incentive for external summative evaluators to give a positive evaluation, because this is their bread and butter. If they are going to get hired the next time, they want the client to be happy.

AF: Right. A client relationship there.

JG: There are pressures there. You can't get away from that. I have those pressures, and they have those pressures. Both. But I do think there is a certain way in which they are more objective, as I said. So they do that.

In terms of external research, we have had researchers who come and basically pursue their own questions. And that can fall on a continuum.

From, we set up a table for them- for example, Alison Gopnik has come over or rather her grad students have come over with their little apparatus and all they need is a table. And we give them a table and chairs. And they essentially use the Exploratorium as a place to recruit children, because schools have become very difficult places to recruit children for research. They have their own lab on campus, but they have to bring kids to that lab. The kids are already here. So we'll do that.

And then in the past, there have also been researchers who have come to the Exploratorium, maybe as a fellow or something, to do research that is of interest to us and them. It's more of a joint thing. That's kind of on a case by case basis, like people approach us, ask us, and we look at what they are doing and say ok. We have a whole policy for outside researchers, where they have to fill out an application form and all that stuff.

So there is some of that, but we do a lot of our own research. Most of the research we are interested in, we do ourselves.

AF: Right. Cool. We just ran through some different work that you do and different types of evaluation. Can you think of an exemplary or favorite summative evaluation project that you can describe?

JG: That an external person did for us? Right?

AF: Yup.

JG: What comes to my mind, of my favorite summative, is the summative we did with Selinda Research Associates. Selinda Research Associates, which is Deb Perry's outfit. Carey Tisdal and Deb Perry together did it. We did a Phase I and a Phase II on APE, and Phase I was in the middle of the project.

AF: Right.

JG: So it was exactly like 'Give us your outsider view on what's going on and help us shape the second part of the project.' And they found something that radically affected the project in a positive way-- just really, deeply helped the project. They do naturalistic inquiry, so what they found was this idea that a big obstacle to prolonged engagement was visiting groups splitting apart. Let's say the two of you come together. If you go to two different exhibits, then if Lisa is playing at her exhibit and Alice, you are at yours, and you get really psyched by something, and

Lisa is starting to get interested in hers. But you get really excited and you say 'Lisa, Lisa, come here, come here!'...

AF: Right.

JG: ...What happens is, she gets pulled out of what she's doing and she goes over to your thing. Now maybe you both get into prolonged engagement at yours, but she doesn't get into prolonged engagement at hers.

So they [Selinda Research] were saying 'Can you think of a way to build exhibits so that it keeps the group intact? You've got a lot of one person, two person experiences. Can you build to keep the group intact?' That came up just as we were thinking about large exhibits and how those don't work very well because people interfere with each other. So all of that led to multiple stations, which ended up being the research study we did as part of the project. It ended up being one of the findings that you could hang your hat on and that a lot of practitioners have said 'ok, cool.' And people here too said 'whoa, that's great.' Subsequent projects have used that idea and built exhibits with multiple stations. So that came out of their finding and that happened in the middle of the project... We called it 'summative-phase I.'

...

JG: And it worked the way I had hoped. We got something that really helped us.

AF: Cool. And then what about phase 2?

JG: Phase 2 was helpful, but again, it was much less helpful. It was good. It told us, what they looked at--We came out of that project, retrospectively looking at the exhibits we created and saying, 'do they fall into categories?' And we realized they kind of did. There were sort of four categories: construction exhibits, where people make something; observation exhibits, where people carefully observe; investigation exhibits; and the last one was exploration.

They took our categories and used that as a lens to look at the different kinds of learning and you know, 'were people self-directed?' They used a naturalistic inquiry approach, which is not my approach. And that's part of why I wanted them, I wanted a different methodology. It was essentially gratifying. They found that 'yes, people were being self-directed.' That's what "active" meant for us - people were actively co-creating their experience with the museum. They weren't just recipients of what we were trying to tell them. They [Selinda Research Associates] said, 'yeah, that's what we are seeing and people are spending a long time.' And they [Selinda Research Associates] did a little bit of tracking and timing. We also did some tracking and timing, and found that people were spending more time and so on. It was like 'yes, congratulations, you did it,' which was nice, but to me, it [Phase 2 summative evaluation] didn't add a lot.

AF: shape everything like the first phase.

JG: Right. I think what it does is it gives more to the field. It helps people see, 'ok, there really is something here. The summative evaluators found that there is something too.' And with different methods. We used audio / video, we coded audio / video, and we coded utterances, like, 'do people ask their own questions' and 'do they answer their own questions more at APE exhibits than at classic Exploratorium exhibits?' Whereas they [Selinda Research Associates] used a naturalistic approach. To have both finding that people are asking and answering their own questions more was cool. That's good. It was good in the sort of 'proof' sense. But the first phase was fantastic in the sense of uncovering a nugget about how people engage and what design you might go after in order to foster that kind of engagement.

AF: How do you set up a study so you have a high chance of uncovering that kind of nugget? [JG sighs] I don't know if you can answer that question.

JG: Yeah. Well, I guess I have some ideas about that. One thing we try to do is stay close to practitioners. Because they often stumble upon problems. As one of the developers here once told me, 'building an exhibit is like 10% having the idea and making it, and 90% fixing all the mistakes you make'. [JG and AF laugh.]

AF: That's great.

JG: So, they [practitioners] come across problems a lot. By staying close to them- And that's why I like the structure of this department as a research and evaluation department, where we do a ton of formative evaluation because formative evaluation can often lead you to a question about 'wait a second, is this more than just this exhibit, is something more general going on here? Let me try bringing three exhibits into this. Or let me now try to operationalize this more carefully. Or let me define my terms here and let's try to build something like a label or an exhibit in a way that really fits this definition and then test that because I think something is going on over here with this exhibit.'

So I gave you guys an example of a study that happened in the APE project. I sent you that paper with the labels. And that came out of a conversation in the team, where somebody said, 'Hey, I went to this museum and I saw that they asked this question at the exhibits and it was cool.' And somebody else on the team, an old timer who has been here awhile said, 'we never ask questions in our labels. We don't do that at the Exploratorium.' Everybody asked, 'oh, really, why?' "Frank always said, 'we don't want to make visitors feel stupid.'" Which is absolutely like- I'm down with that, we don't want to make visitors feel stupid, yes, but there are questions that can make people feel stupid and there are questions that maybe don't. So we started talking about it as a group. Asking somebody 'why did that happen' might make them feel stupid, or unknowledgeable. Asking 'what would happen if you did this,' you just try it. So that got us into this whole thing about questions. Then, 'how do you do the question? How do you ask it? Do you support it with a suggestion or not?' And then we did that study.

Another one like that was: One of the developers came up to me and he said ‘I made this exhibit, and I really feel like it is an APE exhibit. I see people spending time at it, they are doing careful observation, I really like it. But what I keep overhearing is like, ‘mommy I have to go to the bathroom.’ ‘Ok, let's go.’ And they leave the exhibit. Or like, ‘I'm hungry, I want to go.’ Or ‘come on, there is a ton more to see.’ And we were talking about it, I'm listening to him and I said to him, ‘but, wait a second, isn't that what you want?’ And he said, ‘what do you mean?’ And I said, ‘well, all the things you just said had nothing to do with the exhibit. They're not leaving because of the exhibit, they are leaving because they have to go to the bathroom or they have to eat or they want to see the whole museum.’

AF: Right.

JG: These are things that are extrinsic to your exhibit experience. If your exhibit was problematic wouldn't we expect people to be like, ‘I'm done’ or like ‘ok, I'm bored’ or that kind of stuff, something intrinsic to the exhibit? And that was a wonderful realization for both of us because then what we started doing as a measure of APE exhibits was, we started interviewing people as they walked away from an APE exhibit or from a regular Exploratorium exhibit, and we asked them ‘why did you leave that exhibit?’ We would say, ‘we know you have to leave the exhibit at some point because we close at 5 [AF and LP laugh], but why did you happen to leave right at that moment?’ And then we categorized visitors' responses as ‘extrinsic’ like ‘had to go to the bathroom’, ‘wanted to go to eat,’ or intrinsic, or both. And we found a huge difference; it was like 60% extrinsic reasons at APE and 40% extrinsic reasons at the other kinds of exhibits. A big difference.

So that “problem” that he was grappling with, like he was thinking ‘what's wrong with it?’ became a measure that led to a result and it became a way of assessing “APE-iness.” The exhibit itself doesn't say you're done. The exhibit offers enough that visitors can just be there as long as they want. But of course, you know, after a certain point, people want to do something else. So, keeping close to them and doing evaluation with an eye out for opportunistic research. That's a way of finding nuggets.

AF: That's great. That's awesome. So the APE summative evaluation you really liked had two phases. I heard you say you did some tracking and timing and we noticed that in the acknowledgments. Can you talk a little bit about what your role was on the summative evaluation piece?

JG: Ok, for that one, they [Selinda Research Associates] did their own tracking and we did our own tracking. Partly that was because we had a really specific question that we wanted to ask. One of the things that we found through our video of APE exhibits and regular exhibits was that people were spending 3 times longer at the APE exhibits. So one of the questions was, ‘well, maybe it's just that APE exhibits attract a certain kind of person. Maybe there are people who want to explore and investigate for a while. And then there are people who just want a cool,

funny, surprising thing to happen and then move on.' And so the way to get at that is to watch the same person go to an APE exhibit and a regular exhibit. So that's what we did. We set up a bunch of APE exhibits amongst regular Exploratorium exhibits that are also good, strong experiences. We tracked people through. I think we tracked like around 100 visitors. And we only included the 80 something people that actually went to at least one APE and at least one regular exhibit. And again we found a difference. It wasn't 3, it was like 2.5 times or something, more time spent at APE exhibit than at the regular.

But I will say that in general, one of the ways we often interact with summative evaluators is, we often provide- well, a couple things-

One, we liaise between them and the project team. So we will work with the summative evaluator and make sure they really understand the project, that we really understand their methods. We don't have control over their methods typically though. In fact, there have even been times when a summative evaluator has pushed back on us and said like, 'hey, this is the way I want to do it. This is how I do my evaluation. Let me do it.' And we said, 'ok, yeah, you're right. Cool.' So they really do have autonomy. But we talk to them. We act as colleagues. Like if you were at VSA talking to someone about 'here's a study I'm about to do' and you say 'oh, well did you ever think about doing this?' They'll say like 'cool, good idea.' So we do that.

And the other thing is, we will often provide data collection staff. So we have a pool of people that are here in an on-call capacity because we have an ebb and flow of work. Some weeks there's nothing to evaluate and other weeks there's a ton. So we have this pool of folks who are ready. We give them a couple weeks notice and they are ready to go.

AF: Cool.

JG: So we'll often allow the summative evaluator to go to that pool of folks and say 'hey, I'm doing this summative, it's gonna take 3 weeks, who wants to work with me?' And then they train them.

AF: Hm. Interesting. Ok, thank you. Let's transition a little. We talked about APE as example of a summative you really liked. Let's transition into talking about GIVE. Compare your role on a project like that as well as the different purposes.

JG: So there was no summative evaluator for GIVE. That was in the days before NSF decided that even research projects should have an evaluator, which I still frankly do not understand. I mean we've now submitted several grant proposals and things are looking pretty good for one. That's a research proposal and needs to have an evaluator. And the way we're doing it and what I've seen other people do in the field is, we basically have that evaluator as an advisor, sort of an advisory role. Cuz you're doing a research study.

AF: Yeah, I don't understand that either.

JG: What is it they are they going to evaluate? They're not talking to visitors. You're not doing something with visitors, or if you are [doing something with visitors], then you're studying it, so what does it mean?

AF: Right, you're already studying it.

JG: It doesn't make a lot of sense. It would have been weird for us to do an hour plus with visitors for GIVE and then [tell them], 'now can you go into the third room and talk to our evaluator?' No. And our evaluator, it would have taken two years to collect data just like us, there's not some exhibition.

AF: Right, right.

JG: So anyway. This whole evaluation of research projects is hard to understand. Especially because the expectation is that we are going to publish in a peer reviewed journal and that's where you get peer review. So why have a person that you are going to pay to be your peer reviewer? Essentially, that's what you're doing, cuz now you're introducing that bias, you know? Why not let the peer review process speak for itself? I suppose the advisor could help things from going off the tracks in the middle of the project. But is that a big problem? Like does that happen a lot at NSF? I have no idea. But that is a question I would have for NSF. What problem are we solving here?

AF: Yeah.

JG: So GIVE had no summative evaluation, it was a research project. Sue Allen started it, she was the PI, she wrote the proposal. I was a Senior Researcher on the project. She and I had worked for Barbara White in graduate school, we were both steeped in inquiry-based learning. And this project was about, 'ok, APE was great, we came up with some cool design principles,' but even in APE the average holding time only went from 1 minute to 3 minutes. It tripled, but still, the average was at 3 minutes, the range went all the way up to an hour. We did have people on video who spent 59 minutes.

AF: That's amazing.

JG: I know, it's cool. But still, we were like, 'can we take it even further? Can we deepen the inquiry even more?' Some of the thinking behind the project was, 'do families even know how to do inquiry?' Here we offer them this APE exhibit, but do they know how to really make full use of it? Both in terms of doing inquiry and in terms of learning together as a group, as a family unit.

And we were also interested in field trip groups. How do field trip students learn together? Especially when you have this chaperone who is a parent and is with 10 kids and only probably knows their own kid and maybe one or two others. So, do they know how to learn together at an exhibit?

So it was this idea, 'could we train people on how to do inquiry at exhibits?' So that was the driving question and we set about thinking, 'what would such a little training look like?' And we did a bunch of formative trials. We hired Suzy Loper...She is awesome! I would give a lot to have her come back to Exploratorium. She is at Lawrence Hall. I adore her both just as a human being and as a researcher. She's brilliant, and as an educator, she's phenomenal.

We worked together to mock up, 'what could this training look like?' We did a big lit review and looked at what have people done in classrooms. So we tried one thing, where it was roles, like, 'you're the question asker, you're the experimenter, I'm the reflector', that kind of thing. We had necklaces you would wear. We tried a lot of different things. Most things didn't work. A big part of what didn't work was having it be too complicated. Because you're in a noisy, on the floor environment, visitors are distracted. What we realized after 6 months is 'it's got to be simple, really, really, simple.' And we also came across the *visual thinking strategies* work, I don't know if you are familiar with that, but it's basically inquiry in art museums. Brilliant. Abigail Housen and Phil Yenawine developed this approach, VTS. Totally just brilliant. You should look at it. [See <http://www.vtshome.org/>]

JG: We went over to SF MOMA, because they were doing it. They gave us a workshop, we were the students and it was very simple. So we started to think about how to incorporate that. We met with people here at the Exploratorium who do a lot of inquiry, explainers, and the Institute for Inquiry folks. We just tried to distill, distill, distill, and just get it really simple.

We ended up with this really simple idea. We're going to help people learn how to ask a good question or propose a good thing to try. And we're gonna really push on people sharing what they saw and observed. And we figured the middle step of experimenting kind of takes care of itself because the exhibit is right there. What else are you going to do?

So we were torn between how structured do we make that? We basically developed Juicy Question, which is very structured, like, 'we're all going to stop, we're going to brainstorm a question, each of us, we're going to choose one, then we're going to do the experiment, then we're going to stop again, and we're gonna go around and share our discoveries.' It was very 'schooly.' That's what we realized. We tried it on the floor and it seemed like it was working pretty well. And we thought, 'Ew, this is really schooly, this doesn't really fit with the informal atmosphere of the museum, so can we make a version that is much more spontaneous, but has

the same key pieces of getting people to come up with a good idea and getting people to share it?' And that's where we came up with Hands Off. That actually came from the Explainers. They told us one of the things they do at exhibits is, they say, 'all right everybody, hands off, hands off the exhibit, hands off, one person gets to try it at a time.' So we thought 'ooh, 'Hands Off', catchy, makes sense, cool.' So we came up with Hands Off where anytime anyone has something they want to try or a discovery to share, they shout 'hands off', everyone takes their hands off and then either proposes something or shares something. So it's the same key skills, but a different pedagogy essentially.

So now we have two conditions, two treatment conditions. Then we started thinking about control conditions. The obvious one is just to let people use exhibits and not do anything with them. But we realized, 'wait a second, there has to be an effect of Suzy teaching you Hands Off and being this like super smiley, kind, smart educator, warm. There could easily be an effect of that. It could be if you are comparing people who learn a game, whether it is Hands Off or Juicy Question, to people who get nothing of course you are going to see an effect. So we added this second control condition as well, which we called the Exhibit Tour Condition where Suzy tells you about this exhibit, explains a little bit about the content, tells you a little behind the scenes stuff like 'this was developed by Jessica Strick and she is interested in this.' So you get a little of the history of it. And that, we figured would control for this effect of educator.

AF: Explainer. Uh-huh.

JG: Yeah. But she doesn't do any inquiry with them; she doesn't get them to ask any questions, whatever. That's how we came to this four condition design.

Sue and I are both very interested in experimental, quasi-experimental designs. That's how we were trained in Barbara White's lab. You can infer a lot about causation when you do an experimental design, so we did an experimental design. And this is a little bit tricky because we had to set it up so we could randomly assign people without the educator knowing until the assignment was made. So the tricky part was, Suzy- we had three educators, Suzy was one of them, but I'll just stick with Suzy.

AF: Ok.

JG: Suzy not only taught whatever it was she was doing, the tour, the games, whatever, but she also recruited the family groups. So she would go out on the floor and she would do a quick filter interview, like 'do you have the right age kids etc., and are you willing to spend an hour with us?' She would do this quick thing and then she would bring them back to the lab, which was at the back of the museum. It was really important to us that she didn't know which condition she was about to do, because we didn't want her recruitment to be tainted by condition. We didn't want

her to be like, 'let's say I'm doing the pure control condition where I don't interact with people, well this family is fine because they don't look like people I want to deal with anyways.'

AF: Right. Yeah.

JG: We didn't want her thinking that way. So she would go out and recruit people. And we had a person, Adam Klinger, who was both our tech guy and our interviewer. He would man all the cameras from another room. He would make sure the cameras were pointed, were on, and the audio was switched, because every time people went from one exhibit to the next, we had to switch cameras and microphones. So when she [Suzy] left, he would do the random generator, then take a piece of paper, walk it into the room and leave it there. So when she came back, she found a thing that said Juicy Question or Hands Off or whatever. So then she knew what she had to do and then she would do it.

AF: Oh wow, okay.

JG: And she actually said at the time it was kind of fun that way because it kept it fresh, it was like, 'I don't know what I'm doing next.'

We had this whole block design because we had multiple teachers. So we blocked for teacher, meaning we would cycle. Each teacher would cycle through the four conditions, in a random manner, but still they would cycle through them. So we ended up with randomized block design. And it worked really well. I don't know what else, what would you, where you want me to go?

AF: That was good. I was going to ask you about your process for the design and methods behind GIVE because there was a lot going on. You are doing that. Yeah.

JG: We talked a lot about how to actually do it with the exhibits. We knew we wanted to do pre / post and we knew we wanted to do some kind of transfer. It's not just that they can play the game with the educator there and it looks great. We wanted to say, 'ok, but can they now play it at the next exhibit?' Because we wanted a generalizable game, a game they could play at any exhibit. We knew we wanted them to learn it at an exhibit and then try it out at another exhibit and then we would see how they did. So then we thought, 'well, let's have a pre-exhibit where we video tape them to see what kind of inquiry they do at an exhibit and a post exhibit where they've learned the game and then an exhibit in the middle where they learn it with the educator.' So the post exhibit, the educator isn't there anymore, they are just doing it. One thing we figured out quickly was they need more than one exhibit with the educator. So let's teach them the game at an exhibit and then practice the game at an exhibit with the educator there. So we had two exhibits in the middle, but in between the pre and the post, if that makes sense.

AF: Yup.

JG: So we have four exhibits all together.

And there were a couple things we hashed out, that I remember clearly.

One was when people leave that third exhibit, and the educator is like 'ok, I want you guys to use that next one by yourselves.' What does the educator say? Does the educator say 'however you want' or do they say 'I want you to play Juicy Question?' or 'I want you to play Hands Off?' What we decided, and we talked to our advisors about this too, 'let's tell them to play Juicy Question or play Hands Off.' There were a couple reasons for that.

One is that null results are ubiquitous. One of our advisors even said, 'look, sometimes we'll teach kids for 13 weeks how to do inquiry and they still can't do it. You guys are spending 20 minutes with people, 10 minutes at the 2nd exhibit, 10 minutes at the third. To tell them to do the inquiry, they could easily still not be able to do it.'

The second argument in favor of telling people was, it is too weird if we don't. We've just taught you Juicy Question, we've just practiced at the next exhibit. Are we really going to say 'go ahead and use the exhibit, the next one, however you want'? There is all this implied pressure that you would do Juicy Question. Just from a visitor experience standpoint, but also from a research standpoint. There is a confound because you don't know- 'did this family, are they sort of tuned in enough socially to realize that what we really want is for them to do Juicy Question or are they oblivious or are they rebellious?' You don't know. By leaving it open you don't know, so we decided, 'let's tell people.'

The other thing that was a point of discussion was, 'should we randomly assign order? Should we change the order of the exhibits?' Right now it is exhibits 1, 2, 3 and 4 and everybody uses them in that order, but should it be that ok, now for this group, we are going to do 2, 4, 1, 3. Because randomized sequence would be better. But we decided not to do that, we decided to do a straight sequence. And I've gotten this question in talks. 'Why?' There were a couple of reasons. One, it was already complex as hell. For the data collector and for the educator, like I told you, she didn't even know what condition she was doing until she walked back into the room. So then to also not know what exhibit she was going to start with, it just seemed like we were begging for human error. And of course you have to block by teacher for that random assignment, so it just seemed like, and we'd have to block for conditions. We'd have to have an equal number of times when Juicy Question got the 1, 2, 3, 4, as did Hands Off, and equal number when they got 2, 1, 3, 4. You know what I mean?

AF: Yeah.

JG: Oh my god, the complexity just seemed hard to deal with.

AF: Yup.

JG: And then the other piece was, we had these other two control conditions when we were doing a pre / post comparison. We were comparing pre/post change, we weren't comparing absolute. We weren't comparing pre to pre, so it didn't really matter if this first exhibit is an easier exhibit to do inquiry at than the last exhibit, which is the problem. The problem is that the first exhibit could be a great inquiry exhibit and the last one could be terrible or vice versa.

AF: Right.

JG: And it could really affect your results, but the thing is that would affect results across all four conditions. We were looking for change across pre / post so we felt like it's not that big of a problem, let's just swallow it and we won't do a counter-balanced design. So that was another big discussion amongst ourselves on the team. As I said, we still get questions like, 'wouldn't it have been better?' Yeah, it probably would have been better, but man it would have been hard.

The other thing which I didn't mention is that partly, the reason we did it in a lab off the floor was because of the distractions we were seeing on the floor. It was just hard for visitors to really stay focused and for them to spend an hour with us on the museum floor, yelling at the top of their lungs and we wouldn't be able to record as easily. So one of the things we ended up doing in the lab was, we actually took a big table cloth and we covered all of the exhibits and then the educator took the table cloth off for the exhibit they were about to use. Because even just having the other exhibits in the room people would be like 'mommy, can we try that one?' You know? 'No, no, no we have to finish this one.' So again, doing that in a different order every time would be tough. You know? So, yeah, that's where we landed.

AF: That's great, because one of our questions was going to be about two of the challenges you kind of came across and you just talked about several of them, so that was great. *[paraphrases about lab vs. floor]* Can you talk a little bit about the decision to bring it to the lab knowing there were going to be trade-offs.

JG: Yeah. I think we felt like the lab would be an ok place for the research was because the research had particular constraints that the floor wouldn't have. For the research, we needed to be able to see and hear really well. After people did all four exhibits, then one adult and one child from the group were randomly selected and interviewed in the next room and then that same adult and child were interviewed again 3 weeks later. Or maybe it was 6, I think it was 3. But so we did have some measures that were interview measures but really all of our inquiry measures

were going to be audio /video. We knew going in that we're going to look to see 'how many questions did they ask? Or things they proposed? How sophisticated are they?' We knew that going in, and so we just wanted to have really good quality data.

AF: Right, right.

JG: That was the number one. And then the second piece was, we were going to spend an hour with people because of the pre and the post, they wouldn't normally spend. At most you would do two exhibits, right. In the project we did teaching at one exhibit, practice in another. On the floor, that's the most you would do, you'd probably only just do one exhibit, so maybe it's 10 minutes on the floor. So to spend an hour where the kids aren't roaming off and getting distracted and everyone can hear each other really well and the group doesn't break apart. It just seemed like you've gotta do it in a lab. We tried it formatively and it was just tough, it just didn't work that well.

That was a research constraint. That's not what we imagined this thing would be, this game would be when it's really all said and done. So when we did finally bring it out of the floor, boom, all those distraction problems were there, but it didn't matter as much because again, if four people started and one of them peeled off, we don't care, we're just teaching people this thing, we're not trying to study it, we're just offering it.

AF: Right, right.

JG: It's kind of like exhibits, if people stick together, ok great, and if they don't, it's ok, it's all right. Do you want me to talk about what it was like to bring it on to the floor? Is that what you wanted? Because I was telling you about why we decided to bring it into the lab.

AF: Both. I want you to, but I'm thinking, there are so many questions.

JG: The book, the chapter talks about it.

AF: That's what I'm thinking. I do have the book. Let's move on, we can come back to it later if we have time. That's helpful to think through your reasoning for why you brought it into the lab in the first place, what you were kind of considering in your mind.

JG: Sue probably has even more to say about that because she was in the PI role at that time, so she made that decision.

AF: I do want to get to some of your concerns about what we can learn about doing randomized control experiments from GIVE, what can we learn that might or might not be applicable to

evaluations. I know you had some concerns about that, so I wanted to open that up. Is that not specific enough?

JG: I'm cogitating

AF: Ok.

JG: What can we learn from GIVE?

AF: From GIVE. Or you can think about it more broadly, in what ways are randomized controlled experiments appropriate and feasible for summative evaluations or not?

JG: One thing we've talked about before, which I'll reiterate, because I don't know if it was recorded or not. In order to do randomized controlled design you have to have multiple conditions. So, now it's true that one of those conditions could just be a control condition. You could just do one treatment and one control, but typically when you are in evaluative mode you just have the thing that you're trying to make better or that you are trying to assess its quality. In research you might have- like in GIVE we had two kinds of games and that's part of what made it so interesting and what is so great.

This is a quick tangent. Whenever Sue or I give a talk on this, we will ask the audience to make a prediction. 'So who thinks Hands Off is going to outperform Juicy Question and who thinks Juicy Question is going to outperform Hands Off?' And then we ask for a few people to say why. Because there are great reasons why Hands Off should have been better and there are great reasons why Juicy Question should be better. And that's what makes it interesting, because then you really want to know, 'what's the result?' That's how we felt. Half the team totally thought Hands Off was going to outperform Juicy Question, because it's not so schooly.

So one of the problems with evaluation is it's hard to do comparisons, because you don't have different versions. Or it is costly to make different versions, so you have to convince a developer or curriculum designer, or teacher, whoever is offering the thing that you are evaluating- you have to convince them to do it in different ways, even if it means some kind of a control way.

That is another piece. A lot of times I've been struggling with a group of practitioners here, they don't want to have a control condition that is, what they believe in, stripped away. They don't want to do facilitation without all the good things they do.

AF: Right, right, right.

JG: They don't want to do standard whatever because they don't believe in it. So it's the age old problem of giving a placebo pill to people. Is it ethical to give people something that's not going

to help them? And you have to have that whole discussion about, 'what's so important the research in pharmaceuticals is that then once we know, many millions of people can be helped by this pill. If we didn't have a placebo condition, we would have never-' So you have to have that conversation. It's convincing people of that, it makes it hard. This is why randomized is harder.

And then another piece that makes an experimental random design so much harder in this environment than quasi-experimental is that it is a free choice environment. So it's often really tough to set it up so that you get to tell people which thing they do, when. So in GIVE we would bring people back to a lab, so we could decide on their walk back if they were going to get Juicy Question or Hands Off. But with exhibit stuff that's on the floor, a lot of times, you put out this version, and then you put out this version, or you have both versions out at the same time. But then you're not randomly deciding, someone just goes over to this one and someone goes over to that one. So there is a confound with attraction, like I'm attracted to this one and you are attracted to that one. Now there is a confound there. Do I know that you and I are similar enough in all these ways to call it similar groups?

AF: Right.

JG: So it's tough.

I've done it with that label one with the questions, where I had the exhibit with three labels out and I covered up two of the labels and I had one label up. And I grabbed a visitor and said, 'Would you agree to be interviewed for this?' 'Yes.' So it was a cued situation. They would do it and then I showed them the other two labels. And then I randomly selected, what's the next label going to be? I just showed label A and now I'm going to start with label B. And I'd slap label B up and grab the next person. So it's possible, but you have to have a situation where you really can have different conditions and you have control over what the next person gets. Were either of you at that talk about the art museum that started up? You were at that talk.

LP: yeah.

JG: So amazing the natural experiment where they did randomized assignment.

AF: Yes, this is the one in Arkansas. Yeah, we wanted to talk to them. *[JG asks if AF was there at the VSA talk; AF says no but she'd read something about it and LP was there at the VSA talk.]*

JG: So it was amazing. What a great set up. They know who these schools are; they can randomly assign them to come this year or next year, which is treatment and control. It was

great. That's rare, that's a rare find. I asked him, 'How do you keep your eye out for these natural experiments? This is awesome, give us some tips man. This is cool.'

AF: Yeah, yeah.

JG: I think that's part of what makes truly experimental design difficult.

Quasi-experimental design is not that hard though. We've done a lot of studies where we'll have one kind of exhibit, like multiple station version of the exhibit out for a day or a week and then the single station version of the exhibit out for a week and we collect data, then back to multiple station. But I'm not randomly assigning. That week, whoever walked into the museum, they are going to use one condition and then next week whoever walked into the museum, they are going to use the other condition. I didn't get to randomly assign those people, so it's hard to be sure that they are similar groups.

AF: Right, there's confounding...

JG: If you happen to get some bus tour of physicists one week, you are in trouble. So you try to make sure you collect data on the same days of the week, and the same hours of the day and things like that, but you don't know. You just have a large enough sample and then you just hope. And for a lot of our stuff, we don't even want to interact with visitors, so we don't even demographically survey them and say like, okay, 'who are you?' You can do that after they use it, but that is costly and difficult. It's tough.

AF: So I heard a couple of key things that you said make it hard to do. One is the ability to find people or develop or have the time and money to create comparable conditions so it's not just, you get this or you don't get it, you get two versions of something. The other is being able to control what people do and when they do it. That's why randomization is so hard; you have to exert control over people's experience, and it's hard to find situations where that's possible.

JG: Exactly. So, one of the ways we get around the first problem is that when we write grants- and NSF has made this much easier in the last couple years- we will put a research project inside of a exhibit development project.

So for example, the Geometry Playground project where we developed an exhibition on geometry using large scale immersive exhibits as well as table tops. Our question was, 'what does immersivity do for people in terms of engagement, spatial reasoning, interest?' We have a suite of things we were interested in. And because we had written in a 600,000 dollar research project inside of this multi-million dollar development project, we could say to the developers, 'ok we made this large scale thing, it's really cool. Can you make me a small scale version of it?'

Or 'we made this table top that is really cool, is there a large scale version we could create?' So the exhibit is the same, it's the same content, but it is at different-sized scales. And then we can study it. But even there we didn't randomly assign, we had one condition that was all table tops and that was out for a week, then we had one condition that was all immersives, and then we had one condition that was mixed. And we recruited people for each condition, and we cycled through the conditions. So it's a week, a week, a week and then a week, a week, a week. We did even randomly change the order of which weeks, but still, for that week you get who you get. There is no random selection.

AF: Cool. Thank you. We've talked about it in passing. I'm glad we got to explore it with a little more time. One of the things you mentioned is this is a free choice experience.

JG: Oh, wait, so you just mentioned two things, you reflected back two things- it's the resources and then there's control over, cuz it's a free choice environment. And I think there is a third one, which is related to the resources, which is just convincing practitioners that it is worthwhile...

AF: Oh yes, you did mention that.

JG: ...because we are dependent on them usually to make the conditions. With the labels thing, I was able to write those myself. I had the writers help, but I could do that. If it was changing an exhibit, I'd really need them.

AF: Yup.

JG: In terms of really doing something that I would stand behind in terms of research, I need a developer. And that's partly resources, but it's also partly like convincing, 'this is a worthwhile question, this is a worthwhile way to do it.' Having their buy-in.

AF: Yup. So they can help you develop that thing you want to compare or to understand the value of doing that comparison in the first place?

JG: Right, and a third reason, which is again, going back to keeping close to them is, if I want to study designs that are actually going to get utilized in the field, stick close to the practitioners, because if they say, 'well, I would never do it this way.' Then that's not good. I shouldn't study that. I want to study what they think is a strong way to do it.

AF: That's great. If you think of more things, feel free to jump in and interrupt. I was saying you had mentioned the free choice thing, that's something we've been wrestling with as we've been reading and talking to people. But it is a real challenge in informal science settings, this question

of being authentic to the free choice experience while meeting the research or evaluation needs that you have. What are your thoughts on that issue generally?

JG: Wait, just say it again, I'm a little unclear on what the issue is.

AF: Sorry, the challenge of meeting your evaluation or research needs while allowing authenticity of the free choice experience for the visitors.

JG: Right. This is one of the big problems with work in this- in any arena is reactivity. And, um, and it's probably a bigger problem here than say, in a classroom, because here, where people go and how they choose to spend their time is a big part of the experience. Whereas in a classroom, kids essentially do what they're told. That is an oversimplification, but- So if you start saying to people 'I'm watching you right now,' that can have an effect on what they do and how long they spend doing it. Certainly in the GIVE project, they were in a lab, there is a camera right here [gestures] and there were microphones right here [gestures]. It was very obvious we were recording you. [AF laughs] So one of the ways we deal with that is that in our comparative studies, we just do the same thing to the control condition folks. So we just basically say, 'yes there is going to be reactivity, but if we are comparing treatment to control, then everybody is reactive.' And we're hoping that everybody is equally reactive. So that is part of it. Thinking to yourself and in your design, making sure that with the treatment and control [conditions], any reactivity that happens will play out in a similar way, or it will be of a similar quality so that they come out in the wash. So when you make your comparison on whatever measures you are doing, that reactivity comes out in the wash.

The other thing that we do a lot of is uncued interviews, where we let people walk up to an exhibit, use it, and then we'll interview them afterwards. So then you get to see what they would do normally before you mess with them. Another thing we do is we have this, essentially waiver of consent, implied consent procedure, where we have the posted signs and people go into the space. Now the signs are really obvious and there are signs on every exhibit and we know from interviewing people that the vast majority, when they leave, know they are being recorded. But our hope is that when a person isn't there, there is less reactivity.

There is actually hot off the press findings about this from my colleague, Scott Pattison who is doing a study right now, where he's comparing my set up to someone stopping you and having you sign a consent form. And he's finding that people are spending longer, significantly longer, when they sign a consent form than when they do the implied consent.

AF: Interesting.

JG: He hasn't published this yet, this is just personal communication. What he has told me so far is that it is looking like there really is less reactivity with that approach, which is great. That was our argument years ago to the IRB, it was, 'it's a free choice environment, we're trying to interfere with the experience as little as possible. Could we do this posted sign?' They said, 'yeah.' So that's one way to deal with it.

AF: We've run through a lot of issues and things that you think through when you are choosing designs and methods and how you work through some of the challenges of doing, particularly, randomized experiments. Is there anything else you'd like to share about how you select and develop appropriate designs and methods?

JG: Before we go there, let me just go back to the prior question.

AF: Yeah, I'm sorry.

JG: I just thought about this. I don't know if this is part of what you are asking, but because we work in a museum, we're not academics, the visitor's experience in the museum is of paramount importance. So we try really hard not to do anything that will be a detriment to that experience. That's just part of what gets factored into how we do our studies and what we study and what we'll do. Even the GIVE one, having people spend an hour of their time, when they are only in the museum, on average, for three hours, felt like a big thing, so we gave out gifts. I'm talking like, a really nice \$30 gift from the store, which you're not allowed to do anymore apparently, with NSF.

AF: Oh really?

JG: You can give cash, but you can't give gifts.

AF: That sounds strange.

JG: It is really strange, and I think it is actually problematic. Cuz if you read the social psych literature and behavioral econ literature, cash is very different from gifts and it actually moves people out of a frame of gift exchange into a frame of transaction. So if you were to say to someone 'thank you so much for your time, here is a small token of our appreciation.' And it's a little present, \$5 thing from the store, they go 'thanks'. [JG uses upbeat tone.] But if you say 'here is \$5 in cash' it's like 'I just spent 30 minutes and you're only paying me 5 bucks? I make \$100 an hour dude, whatever.' And so all of a sudden it becomes-- We need to deal with that. We're not sure how we're going to deal with that.

AF: That's interesting. Like you just said, there is a clear value you are putting on the time that they just spent with you.

JG: Instead of it being like 'you did me a favor, now I give you something.' It's 'you did something; I'm paying you for it.' And if I'm paying you for it, shouldn't I pay you at a reasonable rate?

AF: Right. I didn't realize that.

JG: It's new. Anyway, we were really unsure about this, 'can we really take an hour of people's time?' And in the interview, we asked them about being in the program and so on. People said it was fine, so it was ok. But when we are doing stuff on the floor, we try to keep it short. We try to have it be experiences that they would enjoy. One of the ways that we made it feel better to them, and this was true, we said, 'you are going to get to do something that other visitors don't get to do.' They really did, especially in the treatment conditions. I feel like, they got value added, they learned, we did offer them, 'here's a way for you to learn together.' And even in the control conditions, they got to use exhibits that were not on the floor, they were off the floor, so there was a little bit of sort of, cachet.

Sorry, coming back to anything else in general. Was it on how we choose?

AF: On how you go about selecting and developing appropriate and useful designs and methods.

JG: Yeah, well in terms of choosing research questions, I already mentioned sticking close to the practitioners. Doing formative evaluation can often lead to good questions. Also writing grant proposals, having us be involved in the creation of a grant proposal that is for developing an exhibition. We will be asking the practitioners who are behind that proposal, 'what is it you're hoping, what kind of learning are you hoping to engender here? Why are we doing this? What is the point of this exhibition? What are you hoping for?' And then really trying to dig down deep in there and say, okay, 'is there a good research question here?'

In terms of methods, it's hard to say. The methods have to match the question and the goals. We have developed a bunch of different methods. Several of us are very interested in methodological development. So, Joyce Ma has done amazing work in coming up with methods for having people tell her their emotional state throughout the visit. It looks like we are getting an EAGER [NSF EAGER: EARly-concept Grants for Exploratory Research: see [http://www.nsf.gov/geo/plr/opp\\_advisory/briefings/may2010/gpg\\_rapid\\_eager.pdf](http://www.nsf.gov/geo/plr/opp_advisory/briefings/may2010/gpg_rapid_eager.pdf)] grant to develop methods for using mobile devices to gather data on visitors, where visitors own mobile device would gather data. They would opt in and it would gather data on them. We are always

looking out for good methods and we just look to see, 'how can we match the methods to the research question and the goals?'

*[AF and JG talk about additional interview questions, especially uses and stakeholders for research and evaluation--we can do this at our next interview. AF says it was really fun.]*

JG: it's enjoyable to think about this stuff and articulate it.

LP: really interesting. Sue mentioned the Going APE project, so I've been reading that over the past couple of days, so it is really fresh. It's interesting to hear about that project and also the downhill race. It was so digestible, especially after reading all those reports.

*[AF, JG, and LP talk about length of evaluation reports]*

JG: We should come back to it when we're still recording, but length of summative evaluation reports I think is a problem in the field.

AF: We've had to read them, so we're also concerned.

JG: We can't get our practitioners to read them because they are so long. We've even done stuff where we digest it and turn it into an exec summary, that is shorter than the exec summary, and then they're like 'ok' and they'll look at that and we'll have a discussion. I won't name the evaluator, but in a recent summative, we actually wanted to give them feedback before the final draft, and one of our pieces of feedback was 'this is really, really long guys, can you shorten this?' And they basically didn't. It's a 200 page report. I mean, I think my dissertation was 200+ pages, I'm not even sure. Where do these guys have time to write this stuff?

AF: we will come back and talk about this when we talk about uses and stakeholders and communicating to different audiences.

Thank you.

[END INTERVIEW]