

Interview on Evaluation in Informal Science Education: Sue Allen

Suggested Citation: Fu, A. (Interviewer) & Allen, S. (Interviewee). (2013, October 24). *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: October 24, 2013, 11:00am to 12:30pm (Pacific) (interview 2 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 second of two interviews that SK Partners conducted with Sue Allen. Much of the first interview, and part of this 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: Ok, so let's just jump in. Like I said I wanted to spend some time talking about stakeholders and uses, so let's do that. I would love to hear your perspective on who you saw as the primary stakeholders or users of your summative evaluation project.

SA: Yes. So the institution and the individuals involved on the team I think were the most obvious stakeholders.

The PI, the core team expressed a lot of interest in it. The exhibit developers, the designers, the team, were very engaged.

The institution, obviously it was a high profile project and high priority project for them. I think at the time it was one of the few children's museums that had received NSF funding from ISE so there was also the sense that it would be relevant to other children's museums beyond the project's life.

And obviously, you know, the whole premise behind posting these things to informalscience.org is that it would be helpful to the field more broadly. I always try to put methodological detail into my reports so that people can see how I got to what I got to. And I'm a believer in sharing methods, so I was hoping that someone who was doing work with children's museums might be interested in future. And there are the details of the tracking study, which was kind of innovative, and some of the details of the interview process and the use of photos, and things like that. So there were some methodological pieces I thought would be interesting to the field.

And then practitioners are another kind of stakeholder, and also any kind of institution looking to build relationships with ethnic communities that they may not have strong relationships with to begin with, like CDM had done with the Vietnamese community and the Latino community.

So, let's see, who else? I was thinking of concentric circles of possible influence.

AF: Yeah.

SA: I think the only people who were really involved as very active stakeholders, in that they were participating in decisions, were the core team. I did also have these consultants that I brought in for my own professional development and support, but I don't think anyone else was directly involved in shaping the studies other than the evaluation team that I worked with, you know, Adam and Joyce and Mary.

AF: So let me run through though and make sure I have a sound understanding of what you just said.

- So the first one was CDM –and two levels of that—the institution as a whole as well as the core team including the PI and the developers, and the designers. [SA: Yup]
- And then kind of, as you said, concentric circles, kind of thinking more broadly, other children’s museums in the field.
- And you talked about potential audiences being a) researchers or evaluators who would be interested in the methodological details and then b) maybe practitioners who were interested in doing things like building relationships with people in their communities.

SA: And one group I left out was the potential renters of the exhibition, especially when you think about all the recommendations that were quite specific to the people who might take the exhibition.

AF: Ok, that’s great, so I have potential renters of the exhibition. You also mentioned towards the end, people who were actually involved in constructing the project itself. So your evaluation team of Joyce and Adam and Mary. Would you consider them-

SA: I was thinking, you know ‘stakeholders’ is such a vague term. And it can mean, at least for me, it can mean both people who shape it and people who learn from it and some combination of those all along the way.

AF: Ok

SA: So people that you may not even know directly are sort of virtual stakeholders. You have in mind a certain audience when you write something even though you may not have met specific people in that audience who do benefit from the work. But the team, the evaluation team that I worked with did shape it in a very direct way in the sense that I would meet with Adam and say ‘how is this working?’ And he would say, ‘you know it’s just too much to try to capture the level of detail with all of these behaviors, we need to drop some of these.’ And I would say ‘ok, great, let’s drop these.’ And so we piloted a lot of aspects of this, including the evaluation methods and even the analysis methods, running that by CDM team and also the evaluation team.

AF: Ok, great. So what I’m going to ask you is what happened with these different groups and that is where I am going next. OK, so for the CDM team, how were the evaluation findings and recommendations communicated and disseminated to them, with them I guess?

SA: Ok, well, I can’t remember how many I did, but I probably made two presentations to the team, with preliminary results for their response and then final results. And, again that shaped the nature of the final report and whether they wanted me to dig deeper into something or less deep into something else. And there was a set of provisional and final reports that went back and forth

with their comments. And then the final report I gave to them. I gave them copies, and I sent them an electronic version and I posted it to informal.science.org as required by NSF.

AF: Right. I know this was a long time ago so it's ok if you don't know, but I'm wondering if you recall who was at those presentations.

SA: Well for sure it was Jenni Martin and Peggy Monahan. Those were the two main people I felt were particularly interested and invested in the results of the study. And I'm pretty sure the designer, Tina was there. And, let's see, I think someone else from exhibits was there. I can't remember. I think Michael Oppenheimer was there when the project started, but I think he was gone by the time I did the final presentation. There were some male staff on the exhibit side, I think. And then, my sense was that it had been an open invitation for the team to attend, so you always get people coming and going on a particular day. Those were the main people.

AF: OK, that's great, and then similar question about reports—the provisional and final copies—do you have a sense of who those went out to and who you got comments back from?

SA: Ah, well again, I think I probably sent them to Jenni, Peggy and Tina and said, you know, 'spread these to whoever you think should get them.' I wasn't going around and recruiting other eyes on it.

AF: Right

SA: Letting them decide who they wanted to see it. And then of course, knowing it was going to be public and electronic made it easy to not worry about that because anyone was going to be able to find it if they wanted. I'm trying to think who gave me comments, but I don't really remember. I don't think I had sort of detailed line-by-line comments from anyone. I think it was more like Peggy, Tina, Jenni, saying 'this is interesting to us' or 'tell us about that' or 'wow, that's surprising'. Or, things like that. I think they were more giving general overview comments about their reactions.

AF: Ok, great, that's very helpful. So let's move on then. So, once the findings and communications were communicated out in all these different forms, do you have a sense of how they were used by the CDM team? By the project team?

SA: No I don't actually. I have no idea at all. [AF and SA laugh] I know that is quite common. Actually that there is no incentive either for the evaluator or for the PI to cycle back in any kind of formalized way, so that anything you do share or do learn is from running into people at conferences or meetings. Of course, they may have used it in all kinds of ways, I really don't know. It just strikes me how, human nature being what it is, there isn't a structure in place for that. I was thinking about that a lot when I was at NSF. Like, how do we create incentives that

help us learn- help us leverage as much as possible from these enormous investments in evaluation? What are the incentives for all the parties involved?

AF: Yeah, can you say more about that? That's one of the things we've been really grappling with too and one of the reasons we focused initially on summative evaluation is because they are often really big investments and it's an interesting question to think about 'how do you leverage it so that it has this broader impact? So can you talk a little bit about that issue and the incentives you thought through?

SA: Actually I think ISE/AISL is ahead of a lot of NSF funded programs because of the push on evaluation and these reports being made public. That's an enormous contribution to the field, which was championed by Dave Ucko and then of course maintained by CAISE and others. I think it's a really powerful and very wise decision that was made years ago.

And only now are we really beginning, with the BISE project for example, to dig through these in a systematic way. It's such a great way to go in terms of recording and mining that stuff. It's fabulous.

Another issue that comes up a lot is fragmentation because the evaluator you have on a particular project may be completely different from the evaluator on the next project. And if you are taking bids from a range of evaluators, and you can't guarantee when the funding is going to come in, there is this issue of 'we all changed partners in the dance.' We can't just assume that there will be 10 years worth of funding going to a particular evaluation company. At the Exploratorium we liked to bring in a variety of summative evaluators because it helped our staff to get professional development. So I think there are lots of reasons that there is this discontinuity. One way to get continuity would be for me to go back to CDM for their next project and say 'well, so how did it work?' You know? [SA laughs] And for a continuous line of funding there is some incentive for CDM to say 'here is how we built on the last set of- here's what we've built'. But I think even then, NSF is always pushing for innovation, so there is slightly less continuity than if you were just pushing for, say, long-term impact as your main outcome.

AF: Right, right.

SA: And just another thing I would share is something that bothers me a lot and maybe you have some wisdoms on this, the three-way collusion between the evaluator, the PI, and the funder. Everybody is happy if the project was a wild success.

AF: Right.

SA: And so there is this foundational tension between research and evaluation in that way. And actually even research is kind of subject to that bias. The things that get published are the positive results often, not the negatives, and yet we all know that the negative is what you really learn from, or at least part of what you really learn from. And I don't think it's solved by just having independent evaluators or even changing evaluators. There is still this strong bias in an evaluation project to show the effectiveness of interventions. At the same time, NSF has put itself into this bind I think because it is sort of accountable for success and impacting national education, and yet at the same time its primary focus, and definitely what Joan Ferrini-Mundy has brought in since she has been leading EHR, is on knowledge-building and learning. So I don't know how to get away from that tension.

AF: That's interesting.

SA: I was just thinking that at NSF, we had these Committees of Visitors, which was our form of accountability and feedback. These wonderful people would come in every three years and look at our programs and say 'here are our suggestions and recommendations.' And we would go 'oh that's great, that's interesting' and 'thank you very much.' But what stopped it being a static, unused report is that we were accountable, 3 years later, for addressing those recommendations when the people came back again. Even if they are not exactly the same people, one of the key things that they're always going to do is to look at the previous COV report and see what you've implemented. So you have to prepare responses, before they come, to the previous set of recommendations. So that's an example of a really structured way to ensure that the people take the recommendations seriously, and think ahead of time about which ones have they taken, which ones haven't they, and why. And it got us to actually do the things that were low hanging fruit and it got us to think about future planning and it stimulated us to get continuity, to remember what our big-picture goals were. There were a lot of good outcomes, and not a great cost to bringing the team together to do that kind of reflection. So you can create systems that enforce that. And presumably NSF could require that too, or funders in general could require that and say, 'before we're going to fund anything else to this organization, we need to see how you have responded to the last set of summative evaluations.'

AF: Do you know of anyone who does that? I don't, so I'm curious.

SA: I don't think I do. Although there's only so much of the world that I know, but I'm not aware of anyone. And of course it's hard to ask the evaluator to volunteer their time if the project funding has ended a year back.

AF: I can't think of anybody off the top of my head, like any funder that does that—who requires you to respond to the recommendations before funding you again. And that is interesting because

that information, of course, is not typically in an evaluation report. And so it would be interesting to have a more formal process where that is documented.

SA: Wouldn't that be great. If you could just tack it on to the final evaluation form, the one-year later follow-up.

AF: Yeah, exactly, a short little write-up. Yeah, that would be cool. Ok, thank you for going with me off on that little detour.

So, we were talking about the Secrets of Circles project team – the ways you communicated findings to them and then you were saying you don't really know how they were used. Is that fair?

SA: That's right. That's because I moved out of the Bay Area so I lost that immediate physical continuity. Though I have seen Peggy a few times since then. I think there is also this general sense of wanting to keep a good relationship with these folks. And it was a very positive evaluation, so you know, do I really want to spend my time with Peggy pulling up some minor recommendation and saying, 'look, did you write up the marketing stuff in the end or did you not?' You know, without a structure, the social nature of even a professional relationship doesn't really support you looking back at the niggles. I guess the bigger picture recommendations might. But then the bigger picture ones are often beyond the scope of the specific project.

So here is another part of the fragmentation issue, that they are accountable for just this piece. And for me to get very grand and say 'I recommend that CDM try all these other ways of building relationships with the Vietnamese community.' I mean that may be something that they are keen to do, but it's just on the edge of what is even appropriate for a specific project evaluation. So there's another reason that it is hard to get that continuity of 'how have you done in terms of the Vietnamese community?' I think the more you are making the evaluation address the long-term trajectory of an institution, the more you get opportunities to also ask the institution, 'how is it going?' But that's relatively few if you keep to the letter of the scope of work.

AF: Yup. This is interesting. One of the things we've heard from the people we talked to around this project is that they have found your evaluation to be useful for planning and learning at the institutional level. So for example, findings from summative evaluation, people told us, informed the direction of subsequent projects, such as the VADI project, the follow-up work that they did to reach out to the Vietnamese community. And that's an example I think of one of those more long-term initiatives that this actually ended up feeding into.

My question was around, did you aim for this from the outset? Did you know that this was a possibility? And if so, how did you position the evaluation to be useful in that way?

SA: Yeah, there was that last piece in the report about framing for the community involvement, and I suspect that that came up because of the earlier framing from Jenni. I mean it was so clear that they were dedicated, that they had a trajectory as an organization towards inclusiveness in reaching out to various communities. And they were ongoing and had done that in multiple contexts and were very committed to doing that going forward. That was just so clear that it sort of invited me to take that perspective. So I think I was looking for that, things that might conceivably be generalizable and future-looking to the extent that I could do that. So it's great to hear that some of that was helpful, whoo hoo!

AF: Yes, they all said there were nuggets in this evaluation that you did that informed their planning for next round of the VADI project. So yes.

SA: That's wonderful. Oh and of course there is the ASTC newsletter. I don't know if anyone mentioned that to you, but I got this call out of the blue while I was at NSF, and someone saying, 'I ran into your study and I want to write a piece for the ASTC newsletter about it.' In fact it was such a joy because I didn't even have to write anything. I've forgotten her name but she said 'I'll just draw pieces from the report – and I'll write the whole article.' I can't remember the details, but I think it was just a lovely situation where someone saw an opportunity to put it out in the world of the practitioner, which I think the ASTC newsletter has a much better chance than any informalscience.org project, and so it was just ready to have it out there.

AF: Ok, so for the broader field, how were findings and recommendations communicated and disseminated?

SA: There was the informalscience.org and there was the ASTC newsletter, and I'm trying to remember if I did presentations on it. I suspect Jenni would have used it for presentations in her world to children's museum conferences and so on. I do periodically go around and make presentations about various projects that I've worked on and mostly I talk about the Exploratorium, but I may well have talked about CDM as some examples. And I think I did actually offer it in when I was working on the NRC report, Learning Science in Informal Environments, I think I threw in the connection to CDM as an example of a museum that was really working to increase its audience, and probably for SBS, the Surrounded by Science volume as well when they were looking for practical examples to show the various issues. Whenever it feels relevant, I share it as a resource—the institution and the report, and I say 'it's online and these people are doing great work and you may want to contact Jenni.'

AF: Great. Once results were communicated out to the broader field, do you have any sense about how they may have been used by others in the field?

SA: What a great question to ask. I really don't and it's just this crazy thing that we do. We throw these things out, this sort of seed dispersal model, even though we know the seed dispersal model is pretty inefficient and not great. Yeah, isn't that interesting. You know I was just thinking too, I didn't mention this, but I know of at least one evaluator who I think has the view that it's the responsibility of the evaluator to check back with the organization at the end of all of this. Maybe I should be using that model - make it on the evaluator's plate to do a one-year check in with them after the project ends.

AF: That's interesting, instead of coming from the funder.

SA: Right. Although I think it is the funder that is going to drive any of these changes.

AF: Right. Are there other uses or dissemination strategies that we didn't get a chance to talk to that you would like to share.

SA: I think there is probably a whole level happening at the practitioner level. I bet Peggy Monahan goes out to ASTC and talks about the project in different contexts. And she has changed jobs a couple of times since then and I bet she uses it in her job talks and as a benchmark experience for her. And probably Tina too in graphic design. I think for each of these expertises on the project there is a set of networks that are reached through word of mouth, conference, general networking, whatever. And I'm sure Jenni built it in to all her new proposals and probably some of the professional development for her staff. I don't even know, but I would imagine that the PIs on these projects use the information in ways that get out to reviewers and get built into the foundation of new projects and more practitioners and all of that. So I think that there is some continuity, having talked about all the fragmentation, I think there is some continuity as people build projects on top of other projects and talk about them to the field.

And of course it will be on the CAISE website too; the project will be listed there. And there may be discussion on ExhibitFiles. I don't even know if there has been, I don't track that, but there may well be. And the ASTC forums is another place. ExhibitFiles, ASTC forums, CAISE newsletters, I mean, those were 3 that were at least relevant to this situation.

AF: Great.

SA: The other thing was that it contributed to the evaluation. It's funny because now everyone is using tablets and variations of tablets to do tracking and timing. The software is much more readily available now, but I think we were actually at the cutting edge of something, and by we, I mean Joyce Ma. I didn't do any of this but she designed it and Adam used it to the max.

So who knows who may have looked at that study in terms of what kind of time resolution, what kinds of behaviors are measurable. I would hope that a bunch of people have had that opportunity to go and read that report. So I think it had methodological contributions to the evaluation community as well.

Also I think some interesting things came up around 1st generation vs. 2nd generation immigrants and the way they might see their cultural heritage differently and so on. A lot of that was framed by Cecilia's ideas, by the way, and I remember her talking with me about some of the data interpretation when I got to that section. So I think there is also a community there who might use the findings, people in organizations looking to expand their cultural offerings to other communities.

AF: Thank you. I'm just curious, in what ways was this project typical of other evaluation projects you have participated in and what ways was it atypical?

SA: Ok as an evaluator, that is the only museum evaluation project I have done as an external summative evaluation. So my experience is quite limited. But as a team member at the Exploratorium I saw a range of them and it was very interesting to see what they did.

I would say, the things that were common were the basic relationships of an NSF-funded project: the team does in-house formative. For Secrets of Circles they had Maureen Callanan's group be part of that. But I was not involved in the formative. My piece was the summative piece with the expectation that this meets the NSF requirement that you have an external evaluation.

I think, in all cases, both internal and external, my experience was that the teams were embracing of this, not rubbing against it. They recognized it as a real opportunity to learn in ways that could grow their organization so it was a growth experience for practice and research.

And then, the basic structure of the relationship was fairly similar, with meeting the stakeholders early on, of having a scope of work, of having periodic opportunities to come by. In all cases, the external evaluators were not local people, so there was this sort of 'what days are you going to come by and talk to the team?' There was the classic PI structure, that you are ultimately accountable for, and there was participation by the team in design and interpretation of the evaluation. So there was, a lot of similarity. Plus there was a recognition that we wanted to come up with some general principles that would be helpful to the field. There was also some sense of remedial evaluation being possible that we could do, because there was some funding kept aside for correcting things at the very end of the project. And there's the general structure of exhibition evaluations: a set of observations, and a set of interviews, addressing some themes of interest. And then you know, general efficacy kinds of questions.

I think what was different was I think mine was far more detailed. [SA and AF laugh.] Is that true? It certainly was very detailed. [AK and SA laugh]

AF: Yes it was.

SA: I know some of the other work like the APE project was very detailed and some of Beverley Serrell's work is very detailed. But I would say it [Secrets of Circles summative evaluation] was unusually detailed for a regular summative, and I would hope it was methodologically thoughtful.

Let's see, what else was different? I should ask you that question. What do you think was different?

AF: [laughs] One of the things we mentioned last time is it the way the report was written and the way you wove in comparisons to benchmarks like Beverly Serrell's work. So that is unusual, I think, given a lot of the reports we've come across.

SA: Ok. Another thing is that not all evaluators interpret data, which really shocked me. Some just kind of put the data out there. We have worked with evaluators who just say 'here's the results, here's the rankings, here are the times. You go decide if that is good enough.'

So I was constantly thinking 'so what?' about this data. What does that mean and how is that helpful and is that good? Because I really don't like data that doesn't move you forward in some way. And so, I think that bringing interpretation and values to a study is just really a key part of it. I think I at least included- maybe it wouldn't win awards, but I did include that, whenever I could, look for benchmarks, and also possible causes, and what this might mean. Why might one see these results? Even if I didn't have data that pointed to a single cause, I would list several possible ones, which again, not everybody does.

And another thing is I think it was quite creative methodologically. I mean we talked a lot about tracking, but also the integrated use of fully bilingual native speakers, cultural natives in those two communities, was unusually strong. And I give credit to CDM for providing me with access to those people who I wouldn't otherwise have had. For example, interviewing the Vietnamese community leader, Quyen Dinh. That wasn't my idea. I think that was Jenni's idea, and CDM came up with this person [Quyen Dinh], and she was able to reflect on the experience of the Vietnamese families who were invited in specially. She gave me a ton of insight, and I used a lot of her quotes directly. Without her I think it would have been much more like a regular report: 'here's what they said and here's what they did.' Whereas she really was a powerful interpreter.

AF: Interpreter of the data

SA: Yes, of the families' experiences. She was there while the families were there, so she could observe it directly and talk about them. And then, it's not like she went through my list of data tables, but I did raise some questions with her. I said, 'You know, I'm seeing this.' And she said 'oh, well that's because...' So it was a bit of a mix of interpreting my interpretations and interpreting her own observations.

AF: Ok, great. At this point we are going to move away from Secrets of Circles and take a step back and think about evaluation in informal science more broadly.

What do you see as the most critical evaluation-related challenges and issues that need to be addressed in the informal science education field? So, in other words, where does this field need to go?

SA: While I was at NSF I made this list of 7 things I thought were the key challenges for the field and put it in the solicitation. So it was actually in the 2011 solicitation. I still stand by those, and those were for the ISE field, not so much the evaluation field, but there were some evaluation related issues too. The other thing I have is that I kind of sketched out five things that I thought were happening in the field that I thought were really good developments.

AF: Yes, let's talk about those

SA: Ok, So I thought:

1) One was "mining lessons from the field." And I had examples of that being the BISE project and the CAISE white papers. The idea that we are actually looking back at lessons learned. I think that is fabulous and really unusual. I mean how many fields do that? Really important. [AF: Yup.]

2) I think the second one I had was, "extending the field" and that is still a critical need. I think what is happening is that the ISE evaluation field used to be a very narrow field, and now ISE and ISE evaluation are both looking out to other fields as was discussed at that last conference. I think the learning sciences is becoming more mainstream as a connection. The field of science communication and public understanding of science is being brought in and connected with, economic and policy kinds of questions are being done more. And I think about Alan Friedman's white paper about policy, and some assessments around the economics of the ISE field, which I know is internationally of interest, and is still very much, very, very much needed. Much of this is being discussed now.

And then an obvious one is extending the field into the school world. And the whole NGSS development. I was just looking at Cary Sneider's presentation at a recent webinar, and he had 6 or 7 points about the wonderful opportunities that NGSS gives us for connecting schools with the out-of-school and informal world. Everything from the raising of engineering as an equal partner with science to the focus that NGSS has on what students can do, not just on what they know. So NGSS has learning situated in activity in a way that informal science much more is already. And cross-cutting concepts as being great opportunities to mix in and out of school, the idea of research and evaluation efficiencies because you actually have a common set of goals. And also having a common set of organizations like 4H that really go national, and can now really work effectively so both the implementation and the evaluation aspects of ISE could be done more efficiently and more consistently.

So you know, those were all his thoughts, but I think that is a really important place for us to be growing and I think we are. I think we are doing quite well on that score. So that was just the idea of extending the field, and a lot of discussion came up when we were all meeting last time in Washington about hooking with anthropology and sociology and health and marketing. And just that there are so many fields of study that are not yet commonplace in the ISE world and the literatures are different and the language issues and just lack of awareness and lack of time and all those things. There is still incredible fragmentation on the knowledge-building side.

AF: Right.

SA: So I think we are making good progress and I still think there is a lot to do. And I think that is ripe fruit because instead of having VSA meetings be the same people every year- actually I haven't been to them for a few years, so I'm just making this up, but there used to be sort of the same people over and over and it was a lovely small little group, but we need some cross-pollination, so that's an ongoing need. That would also help us to be taken more seriously as a field or in practice and in evaluation and research.

3) Then "integration within the field." And by that I just meant, within the ISE field integrating research and practice. I still think there is a long way we need to get there. The research to practice site is a really nice one. And the claims wiki is another one. And I'm sure there are other things going on. But I think the ongoing question, which is not just ISE, it's really everywhere, it's this huge rift between research and practice, it's just in every field, and continues to be a big deal. And it is interesting to note that the latest AISL solicitation really tries to integrate that even more, so I think that could help too.

4) Then I had a bullet called "building assessment tools," which again is an ongoing need that I think we have made some progress on. I think there's building of specific instruments because our field is just really hard to get valid, reliable instruments that aren't intrusive and

invalid. And an example of that is the Dimensions of Success work that Gil's been doing. I'm sure there are others, but I think instrument building continues to be a major need.

And I think instrument sharing continues to be a major need. The ATIS site is a nice example of that, but I think there could be a lot more sharing, a lot more sharing of instruments with the obvious technical support that it needs, to talk about when to build your own and when to use other people's, what are the issues involved, so people don't just kind of take them and use them blindly, but can use them sensibly.

And then sector-level common instruments, I think that is an area where we've really made big progress thanks to a few committed people. In the afterschool world it is Gil Noam, in the citizen science world it is Rick Bonney and Tina Phillips. And in the museum world, Beverly Serrell was first person that I knew of that really took it on, said 'Ok, let's do it'. I think all of those people get flack for it. But the secret, I think, of their success is that they just went ahead and did it and then put it out there and people did come. So I think at the sector level we have some good common instruments being built.

One issue that's still outstanding related to that is measurement of competencies and the whole badging movement. I think that is an area that we could be doing more to hook up with the formal world. You know I am in a state that's just shifting over to proficiency-based graduation requirements for schools. And they're saying 'what does that mean? What do we do? Proficiency in what? How do we measure it?' So once again we're in this world of 'what does proficiency look like?' And I think that some of the open badging work is really a great opportunity for us to connect informals to formals and think about what proficiencies looks like that could be developed in any setting. Because if we take seriously this 'anywhere, anytime, anyone' learning, lifelong learning, then that means we are really believing in the development of proficiencies over time, space and setting. And let's assess them. This could be the best thing that's ever happened is this opportunity for us to be part of that assessment discussion.

AF: [And what are the challenges around doing that?](#)

SA: Well, it's horrible isn't it? It's really hard. Vera Michalchik will say 'these places are fundamentally about having ambiguous, open-ended experiences, and they support proficiencies that are not only multiple and ambiguous, but defined by the person, you know, self-defined.' So you might go to something for one reason and find you come out learning something entirely different, and maybe that's different again from what the program person or the exhibit developer had in mind. So you're up against, you know, not just logistical issues, but big theoretical issues around assessments.

AF: [Yeah.](#)

SA: And there is still a reason to believe in it because I think there is enough commonality in terms of what people want and what we could give them, in terms of proficiencies, that some of those arguments would fall away if we had something that worked reasonably well. So I think it is sort of parallel to the building assessments, sector-wide assessments. Build something that is actually quite good and post some badges and get a few people to do it and you could suddenly get snowball, I think.

AF: Can I ask what kinds of proficiencies you think are of common interest?

SA: Well, practices from NGSS, the idea of recognizing problems and solving them. This is where I think the engineering is such a powerful cross-over point. Because a lot of the afterschool and museum and informal, and even home science activities have this quality that the learners are creating something and expressing themselves through that. And the whole maker movement is another example. So the idea of having portfolios that show one's ability to identify a problem and solve it in creative ways and iterate it, that is just perfect for out of school.

AF: Something that we've been concerned with too, is that instruments often are not shared or they are not included explicitly in a report and they are hard to find.

SA: Oh yes, while I was at NSF I tried to mandate it. I really did. I tried to get us to say, 'If you are going to take tax-payer dollars you have to put your assessment out there, your instruments. You have to. There is just no other way.' And I couldn't believe it but, no, it was not possible for us because NSF has a whole set of policies around intellectual property. And this comes from all kinds of other fields, probably universities doing science research that have to acknowledge that those fundees have intellectual property for the things they create. And so, in fact I couldn't override that. I would have loved to, but no, we couldn't legally mandate that people publish their instruments.

So I still think it is misguided even for the evaluators to think that that is the best business model for them. Personally I think it's a mistake, but they are entitled and I think the way we go around it as a field is by having, again, commonalities. Look at Beverly, look at Gil, look at Rick and Tina. They are putting out stuff. They are getting groups of people together to say 'how about these instruments?' And the people who want to play will play and that'll move us all forward.

Actually, I know this whole thing of common instruments is very challenging. And it's funny because I was just doing an evaluation last week [SA laughs] and I started realizing I really needed to create a variation on this instrument. And I just thought, 'oh boy, I hope nobody finds out I've done this.' Because I'm always waving the common-instruments flag. But you know, for validity questions, I realized this question wasn't exactly what was going to get at what I needed

to know. So I was right in the thick of this. But I guess I am a believer in the Alan Friedman model of having at least a core set of common questions, a core set of instruments that have got some validity and reliability to them and have been used across some set of programs and that are mostly aligned in terms of their goals.

Also, here's another thing, I think we need to be making it really easy for people to share not just instruments but also data. For example, I am creating an instrument, and I'm using NAEP questions because there is a national calibrated set of NAEP responses. I can compare it by state, by nation, whatever. But for me to get easy access to that summary data has been really challenging. I've called up Larry Suter. I've said 'Larry, could you please help me to understand how to get these results?' I'm pounding on the site and I can't even figure it out. And that's best case, when I already know what NAEP is and know that these results are potentially available. So we need to have CAISE to put out not just what instruments are, but what are the nationally calibrated data available and make it absolutely accessible. Because if I can see that there is already data on that question, my proposals will get stronger, and I will use those assessments, I will refer back. It gives me a way to use a benchmark that is out there for the nation. I mean it's just crazy that we are having to work to dig for these things.

AF: Yeah, that does point to something we have seen too, is this lack of benchmarking. There is a lot of reporting of results and you don't know what that means. I have no context for that.

SA: Lack of benchmarking. Right. That's what's made the fields of genomics and astronomy and these other big-data fields catapult into success and growth, is common constructs and sharing of data via these enormous, really structured, well-defined datasets.

So this, now you're getting at my view of social scientists is that we're not really scientists unless we have common constructs that we point to that are measurable and comparable. I don't think we're scientists, I think we're social evaluators or something else, but I don't think we are scientists unless we're abiding by that. So there is my hard-core pitch as an ex-physicist.

And NSF is now mandating, as you know, when you write in a proposal, you have to specify how the data will be shared. And I know from working there that the intention of the data-sharing mandate is that over time the bar will be raised so that we are going to need to be sharing our data with each other. And I'm all behind it. I think it is exactly what our field needs. So you know, ISE could be a leader in that area of saying 'ok, we're going to start off really early by putting out there easy places to put your data, find your data, and tools to combine it.' What is Jon Miller's data showing? There is an example of someone who has done a whole lot of research on national trends and he's got national samples. They're not perfectly representative, but they're not far off. What can we build on that will speak to his instruments and his findings? And they're not perfect instruments but they get a lot of press just on the grounds that he has all

that data. And NAEP is another one. Places where we have interest-based national surveys should be referenced. And Larry Suter is a fabulous resource for that.

AF: I have to ask this because it's been driving me crazy. You were talking about this new push to share data, but I imagine that there is a lot of fear, potentially a lot of fear around doing that. How do you get around that, I guess?

SA: Again, I think the secret is to do it. I think we are doing exactly the right thing by having sectors within ISE build the instruments, so that within a sector, you do get a lot of commonality of programs. The kinds of outcomes that citizen science projects propose are kind of the same. I think taking on the whole of ISE is too much. I think it could kill it because it is just too broad. But I think doing it within a sector is the right grain size. And you do it just gently, inviting some of the field leaders, because at the sector level you can get the six major people to sit down at a single table and say 'ok, what shall we do here? Here is a draft, what do you think? How can we make it better?' The size of the field and the size of the outcome array is just small enough within a sector that that's a good way to go. Once it is out there, you say, 'this is not perfect. Help us to tweak it, but here is what we are seeing and here are some results.' And you start using it and you get a Beverly Serrell escalation of use of it.

I also do think, I mean I have to respect that people are not electrons and you're always going to have limitations. But I think that our field suffers far more from fragmentation and lack of comparability that cripples our ability to make claims as a field. So I would rather lose some validity and be able to make some powerful general points. Because when you have a single powerful point - look at the Robert Tai study. Everyone and his brother quotes that study. It's everywhere. Or Rena Dorph's study about how few teachers are doing science now in their classrooms. So you come out with a few powerful measures. This is why I am sympathetic to Jon Miller's measures of literacy even though obviously they're flawed. And Larry Suter will say, 'Well, right, that's how you progress: you start with a flawed measure and you make it very public and people get all upset and say 'this is not right, this is not good enough.' And what do they do? They sign up to be at discussions in the future to make it better. And that's how TIMMS and PISA and NAEP and all of those guys came into being, as I understand it. And you end up with something that's not bad. And even those, you know PISA is not perfect either, it never will be, but it is being used across the world. How amazing is that?

AF: I think I heard all 5 items on the list. Is there anything else?

SA:

5) The one I left out was that we still have this issue of "legitimacy." I think it is better and I think there is a lot of opportunity, but we still have that as a major issue. And I think we don't do nearly enough to learn, it's kind of ironic, we don't do nearly enough as a field to learn about

what kinds of evidence and arguments and evaluation results are actually seen as compelling by our key stakeholders, by policy makers and the people we want to affect. So I think that is a major need.

6) Then the other thing I had was just the “community engagement.” I mentioned ExhibitFiles, Museums and the Web, and I think I mentioned the ASTC forums, but I especially want to sing the praises of CAISE. I think it is an incredibly valuable central hub for the field and it’s taken a long time and people’s awareness of it is still growing I’m sure. But the service that they do is just huge, to be a central hub in a very fragmented field. With all these things like the white papers they do, and the forums that they do, and the newsletters that they do, and all of this stuff that includes funding mandates and NSF deadlines, and everything you ever want to know. It really helps to increase the chances that people know where to go to find stuff. And all the cross-referencing that’s going on there.

I mean that was very intentionally done, to set up a hub like that that would really serve the field. One of the main things it does is to disseminate best practices, individual projects, cases, facilitate discussions. It’s very multi-faceted in terms of the way it builds community. And NSF encouraged them to do more of the on-ramps and one-stop shopping. So the idea that there is now part of the website that introduce you as a person who is new. It’s all about thinking about how to make yourself attractive as an on-ramp experience to someone from the outside. So part of it is learning what are the anthropologists doing and the other part is saying, ‘if you are an anthropologist, who might want to work with us, here is our language and here’s some introductory text.’ Then you get partnerships going.

AF: Great. In one of our other interviews, we had somebody talk a lot about this issue of making the case for informal science and needing the kind of evidence that people, such as policy makers, would find compelling and he talked a lot about how data from RCTs are more compelling and can make a stronger case. I guess I am wondering what your take is on that and how appropriate that is.

SA: I suspect it depends a lot on the specific stakeholders. And we certainly know that RCTs are not the most compelling to everybody. Because we know the power of the anecdote: Dennis Bartels, who works a whole lot with Congress, has told me numerous times that it is the powerful anecdote that carries sway. And that it’s Congress seeing activities in their own communities, talking with teachers who say ‘NSF funded programs really got me where I am.’ Or it’s seeing their kids watch Nova on their TVs when they go home at night, so often it’s not actually the dry study that is compelling.

So I think it is a really interesting question, and I bet there is a literature on this. There’s a particular high school in Maine that made a short video of its work in expeditionary learning, and

I've now had 5 people send me this link to this You Tube video of this school – it can suddenly go viral. So I think there are some good questions about what constitutes powerful evidence. This video makes a very compelling case. And especially in our field, aren't we about experiential learning? How is this possible that in a field that is built on experiential learning that we feel we need to make our points with reports? [AF laughs] It seems crazy.

On the other hand, I think there are some quantitative claims that are powerful like the Robert Tai study, because there was a powerful factoid. You know it's almost like the experience of being in museums, it makes me think about what kinds of small, intriguing, personalized facts people take away from museums. Things like 'wow, I see how much 16 liters of water is and that's how much is in my body.' So I think the Robert Tai study is an example of that. Or "the 95% solution," there is another one. There is a provocative fact, I think that may be the way to go – a combination of provocative facts and powerful anecdotes may get us further with policy makers and people who we really want to respect us than RCT's.

AF: That's really interesting. Ok, so related to my question about RCTs, one of the things we were talking about is how people draw lines, or don't draw lines between research and evaluation. I'm wondering if you can offer your quick thoughts on that.

SA: I think it's a continuum. And I think at the extremes you have that the purpose of evaluation is to make better the thing or to evaluate the quality of the thing, and that research is really about creating useful knowledge-building principles that can be used in future. So I do see them as different, but I think at the middle they do just bleed one into the other.

AF: We ask that question because we have heard from people, in *Secrets of Circles* it was the case too. Even though you are doing an evaluation, perhaps maybe with narrower purposes around improvement or looking for evidence of whether you met goals, but there is also a drive to look for more generalizable principles and look for these nuggets of knowledge that can be more widely shared.

SA: Yeeesss! So it's so wonderful talking with you because these are all the passionate issues I have as well.

AF: I know, well I feel the same way, I could go on and on [both laugh].

[END INTERVIEW]