

Interview on Evaluation in Informal Science Education: Beverly Serrell

Suggested Citation: Peterson, L. (Interviewer) & Serrell, B. (Interviewee). (2013, August 29). *Interview on evaluation in informal science education*. [Interview transcript]. Retrieved from <http://www.informalscienceevaluation.org/>

Interviewee: Beverly Serrell, Director, Serrell & Associates

Interviewer: Lisa Peterson, SK Partners

Note-taker: Alice Fu, SK Partners

Date and Time of Interview: August 29, 2013, 9:00am to 10:30am (Pacific)

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.*
- *Italics also indicate names of reports, books, papers, etc.*
- '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.

[BEGIN INTERVIEW]

LP: As we mentioned in our email, we are interested in learning from leaders in the field of informal science education. To warm up, can you spend a few minutes walking us through some personal highlights of your career, especially in relation to informal science education and evaluation?

BS: Ok. Can you give me a little bit more of a leading question? That's so broad.

LP: I know, I know, it's meant to be kind of a broad icebreaker. Is there anything that you feel you would like to highlight from your career or things that have stood out to you? How you got started.

BS: Well, there are a lot of people who were really helpful to me, influenced me and challenged me. It's a long list of people. Probably the top names would be Chan Screven for helping and encouraging me and just always through his life. He was always very good to me and generous. Especially in his later years, he was always very complimentary saying 'You were really doing a good job.' That was very sweet of him. Up until the last time that I knew him, he was just very, very nice, just a very helpful, nice person.

Probably one of the other people on the end of the spectrum for challenging me would be Harris Shettel, who challenges everybody... He's always saying, 'Why doesn't this happen? Why didn't that happen? Why doesn't this person do that?' The first time I ever heard him speak at AAM, he was saying that about the people in the field then, and I was going like, 'Wow, I had no idea there was so much that wasn't being done.' That was 30 years ago. Then I had various run-ins from pleasant to pretty painful through the years, when I was on the board of VSA and he was on the board of VSA and had various opinions about how things were going, and I was summoned to try and talk him into or out of different positions at various times. Last time I saw him, was at AAM in Baltimore. Steve Bitgood is another one of the names I have to recognize....

So Harris wrote a really challenging paper that was important to me in the early days and later when I was working on the *Excellent Judges* project. And it's the paper called *Exhibits: Art form or educational medium?* You can find it in my bibliographies in various places. His challenge is to try and be able to evaluate exhibits. I think I met several of his challenges when we finally put together the Excellent Judges framework. So that would be both a highlight in terms of a person and of an accomplishment that I'd like to cite.

Other people that were important besides Harris, Chan and Steve were Minda Borun. She was in the field before I really got into it and her writing and her speaking was very influential. Patty McNamara was also-, preceded me into the field. Her talks were always very clear and

thoughtful. You met her in Milwaukee.

LP: Yeah.

BS: She's one of these people who's just very quiet and never is pushing her own trumpet and agenda and stuff like that, but if you look back over the things she's written and done, she's just a really solid player.

LP: Yeah, it was a pleasure meeting her.

BS: I hope I remember some of the other people to mention them during our conversation.

Another one of the highlights has been writing, publishing and getting recognition for the *Exhibit Labels* book. That was published in 1996. It was the second labels book that I wrote. I wrote the first one way back in the '80's sometime. That one was called *Making Exhibit Labels*. I may have told you the story about going to the Field Museum to look for the book to see if they had it in their gift shop / bookstore, and they said, 'Oh yeah, I think we have a copy, it's over in the crafts section.' [LP and BS laugh] So like, making exhibit labels, like how you cut out the letters. [LP and BS laugh]

So the second book that I wrote was much less about how you cut out the letters and much more about the philosophy behind writing. Some people have called to attention the fact that it's not really a book about exhibit labels. It's a book about thinking from one end of spectrum, the beginning of the big idea, to other end of the spectrum, summative evaluation and showing the impact of good labels.

So having people come up to me when I attend the Visitor Studies Association meetings, the AAM meetings, and say, 'Oh I've read your book, it was so helpful, it really opened my eyes' especially when young people come up and say that, it's very gratifying, it's very, very gratifying.

Now it's gratifying that people are starting to come up and say 'I've been reading and using the *Judging Exhibitions* book.' That's especially gratifying because that, I think, has the potential to have a broader impact on the field, to get people to really be thinking of how exhibits and programs need to be visitor-centered, need to be looking at it from the viewpoint of where the visitor is coming from and not just thinking about the visitor as a receptacle for all the information and knowledge that you want to pass along.

LP: Ok, great, great. So you've mentioned some people that we've heard of from the field that have influenced you. And it's interesting because you talked about how they had influenced your

career and then coming around full circle and some of the things that you've written and published and then people coming up to you and telling you that they've read your books and have seen some of your work, so that's great. Thank you for that summary, that's really wonderful. And as you think about some of the work that you've done, and more specifically evaluation work that you've done, can you think of an exemplary or favorite evaluation project?

BS: Well one of the first, biggest evaluations that I did, that was really pivotal for my thinking and my product back to the field would be the evaluation of the exhibit, *Darkened Waters*. That was a fabulous exhibition. One of the other people who has been a great colleague, both complimenter and challenger of my work has been Kathy McLean, Kathleen McLean. She and I worked on that exhibition. She was very key in getting it organized, designed and written. There was money, there was NSF money on that. That was the first substantial amount of money to do a really good summative evaluation, so we used a big combination of methods in that. The report, I think, was seminal in terms of both sharing some really good methods and producing a report that could be shared with the broader field and they could understand what it was about.

So many times [a report is] either written as an in-house document that doesn't make much sense to other people, or it's written as sort of a fluffy, uncritical review of how everything was really great and visitors learned everything, and did everything they were supposed to. So the tracking and timing, interviews and some other things that we did, I think represented a really good example of how you can do an evaluation if you have a decent budget, but you can do it in ways that represent different ways of thinking about evaluation, different ways of using input from visitors.

LP: Ok.

BS: Plus, it was a really good exhibit. So many evaluations are done of exhibitions that, in my opinion, are not very good exhibitions. We need to get together and have peers, colleagues in the field, agree that these exhibitions are good exhibitions that we think, as a professional we're looking at them. And now we're going to look at them not just as a professional critique, but we're going to use some evaluation techniques on these good exhibitions and see what kinds of feedback we get from that. So I think it's very difficult to get a collective representation of ways of looking at exhibitions when you're evaluating exhibits that have such a wide variety of quality.

LP: Yeah. So how do you handle that as an evaluator if you go in and maybe you don't think the exhibit is as high quality?

BS: Well, I have my opinions about whatever exhibit I'm looking at, whatever exhibition I'm looking at. And I know that the museum wants to be put in their best light. But I also know that I

have to be honest in terms of what the strengths are and what the weaknesses are of the exhibition, regardless of how good I think it is.

And, um, I think. [Pause.] Have I ever reviewed an exhibition that I really hated? I don't think so. I really don't think so. Um. [Pause.] I've reviewed some that I like better than others, but I think I like most of the ones I end up evaluating.

The paper that I'm working on right now for the BISE project. Did you go to any of the meetings that talked about that? At VSA in Milwaukee?

LP: Yes, yes.

BS: There's one report that I'm including in that, that people will be able to go and look up. It's an example of an exhibit that failed in many, many ways. And you don't see people giving that kind of a report very often. And I think it's exemplary to say, 'Look, we don't go into exhibition design intending to make mistakes, we don't try and make a bad exhibit. Nobody ever does that, but sometimes, it turns out that there were many, many, many missed opportunities for a variety of reasons. And here's an example of one.' That is going to be included in my paper. God knows what will ever happen to the paper. They haven't figured out how to get them published yet or where.

But *Darkened Waters* is one to cite. And I know, [Kris Morrissey] out at Washington State. Lovely woman.... She told me that she uses the *Darkened Waters* evaluation in her classes as an exemplary study.

LP: Oh, oh, great, ok.

BS: So I was excited to hear about that. Beyond that one, I'd have to look at my list of things to pick out that I'd say, 'Pay special attention to this one.' Or I would love it if you would look at them and come back to me and say, 'We looked at this one and we like it for this reason or we have questions about it for these reasons.' It would be great.

LP: Ok. That's a great example and I love the way you said that it was pivotal for your thinking, and how you thought about evaluation kind of being seminal in the way that it was about sharing methods with others and then sharable with the broader field. That's a criticism sometimes, that they are specific about exhibits and they're not as broadly applicable or sharable. What was it about that project that enabled those kinds of things to happen? Was it the funding? Or the-

BS: I think it was definitely the funding that gave us a chance to try out a lot of different approaches and really have the time to think about them and think them out clearly.

Darkened Waters was seminal for what was later my efforts to pull together tracking and timing data from 100 different sources and come up with some trends from that data and share that. That was the book called *Paying Attention* that was published by AAM. In *Paying Attention*, which was 1998...?

LP: I've got it sitting in front of me.

BS: 1998 right. So *Paying Attention* came out in 1998. *Darkened Waters*- what was the date on *Darkened Waters*? [pause]

LP: Oh, I don't have that in front of me.

BS: I don't know, maybe 1992 or something? I can't remember! [laughs] I'll have to look it up. Anyway, it was in the early 90s that we did *Darkened Waters*.

And *Paying Attention* contains a workbook that tells you step-by-step how to do the tracking and timing and how to do the cued questionnaire. Now *Paying Attention* doesn't include any of the data from the questionnaire, but it includes the instrument for how to collect it. And I've continued to use that instrument in many many different settings and many many different studies. And continue to consider it to be a very trustworthy workhorse of an instrument that gets me really valuable data relatively quickly and inexpensively. And I've seen some of the questions that are used on that picked up by other evaluators and used because the wording is just right. It's the right words. And that takes a lot of trial and error to get that. And so I'm proud of that and I'm happy to share it.

Some evaluators do not share their instruments. They feel very proprietary about them. They are protective and don't even want the museums that they work for using them without their permission.

LP: Yeah, that's interesting. There is a lot of talk about common instruments currently and I think you said you had the instrument included in *Paying Attention* and you've used it and others have used it. What does it take to create those instruments that are shareable and valid in more than one setting?

BS: I think it has to be a shared objective. Both the kind of information that you want and the ability to stay at a- [speaks slowly], this is going to be hard to frame in the right way- stay at a level of specificity in your questions and stay at a level of general sort of a-, I always refer to it as a broad brush. It's a broad brush, paintbrush- that you're painting a picture with this broad brush and it doesn't allow you to do fine granular little things, but it does give you the ability to

give a broad stroke of data and to hear from people in a very big question way. But within their answers, it's just filled with, I think, gems. But it's qualitative information. It's not easily quantified. So I think people don't understand the difference between collecting data that is going to be analyzed in a qualitative way and wanting to collect that data qualitatively but then analyze it quantitatively. It's hard to do....

BS: I do not consider any visitor to be a typical visitor. And that's a very hard concept for people to understand. There's no such thing as an average visitor. There's no such thing as a typical visitor. But there is data. You can look at data that can be summarized in numerical terms where the word "average" is used and it's very valid.

But there's different schools of thought that are floating around out there. And people haven't sat down at a table together and talked about it long enough to work out those details.

And I think that the paper that comes closest to being a prerequisite to that sit-down is the paper that's in the book called *In Principle, In Practice*. It's edited by Falk. And it's the chapter written by [pause] many authors including Deborah Perry and Sue Allen. And I think it's the best description of the difference between, kind of an open-ended approach with smaller samples using more case study and qualitative information and a more systematic quantitative approach. And the name of the paper is: *Research in museums: Coping with complexity*. It's chapter 16 in *In Principle, In Practice*. And I would love to see a discussion between people representing both those schools of thoughts, talking about them long enough and thoughtfully enough together to reach consensus on some instruments that work for both of them.

LP: Ok.

BS: Not that they have to be shared in exactly the same situation but that there can be some agreement. People, I think, are territorial. They are proprietary and they are sometimes closed-minded and they're just working with different models in their heads. And I think some of those models are useful and helpful and some of them are harmful....

LP: Yeah, so I like that idea of having that discussion from both schools of thought and seeing if consensus can be reached and maybe having that as a way to move forward past those challenges that you mentioned—the territoriality, the proprietary closed-minded...

BS: Right. And I think the discussion needs to be framed clearly and this article helps clarify which camp you're in and which things really resonate with you positively and which ones you're just kinda like, 'Oh no, I just can't do that because it just doesn't feel right to me.' Or 'I don't want to do it that way because I've tried to do it that way and the results that I got when I

did that were not as useful as when I tried it this way and the results I got were more useful because.’

So you have a real discussion that works. Both from the overview on the philosophical and theoretical point of view and gets it down to feet on the floor, in the gallery, looking at what people are doing, talking to people about what’s going on, to make it really real for how this is going to be used in the field.

LP: Ok, ok, well thank you for that recommendation and discussion around it as well. And speaking of philosophical, I’m going to move to another question that ties a little bit into some of the things you were saying before about exhibits that you see as high quality and evaluating those. One common definition* that we’ve seen of evaluation is, begin quote here, “the systematic investigation of worth or merit of an object.” And that’s the end of the quote. And we’re wondering, what does that word “worth” mean to you?

*The Joint Committee on Standards for Educational Evaluation (1994)

BS: Um, read it again. You’re defining what?

LP: It’s a definition of evaluation and it’s from the Joint Committee on Standards for Education Evaluation and it’s, “the systematic investigation of worth or merit of an object.” And we’re trying to tease apart what that word ‘worth’ means.

BS: And are we substituting the word exhibition for the word object?

LP: You certainly can, yes.

BS: Well I certainly agree with systematic part. I think you can be systematic in whatever camp you’re in but you do need to be systematic.

And worth or merit. [pause] Those are really pretty fuzzy aren’t they?

I guess I would substitute words like value. And then you have to say, ‘what value does this exhibition have to the institution that created it? And to the people that used it?’ And that can be goal-referenced by what the intent was. And it can be goal-free as to what kinds of impacts were witnessed systematically and documented that may not have been intended and may have actually been positive. Or maybe there were some unintended impacts that people would want to think about for the next project, doing something differently.

LP: Ok, great, so you were saying the words ‘worth’ and ‘merit’ are kind of fuzzy and you would substitute the word ‘value’, kind of, ‘What value does the exhibit have to the institution and to the people using it or visiting it?’

BS: Um hum.

LP: Ok, great, and kind of looking at either goals or things that are found outside of the goals that were pre-established for it.

BS: Um hum.

LP: Ok, great, great. And what is the role of an evaluator in investigating that worth, merit or value of an exhibit?

BS: Well the first thing is to find out what their intention was. Ok, here you have this exhibition. I see it. I can see where it begins and I can see where it ends. And I can see all the stuff that’s in it. And I can read all the labels that are in it. What did they hope would happen? Why did they mount the exhibition? What did they hope to gain? What did they hope visitors would do and get out of it? So it’s really doing kind of a download of the reasons for why they did it. What worth, merit, or value did they see coming out of it for themselves and for their visitors?

LP: Um hum.

BS: And then I have to look at what those goals were and say, ‘Well, do I have some methods in my toolkit that would help reveal feedback from visitors, either by their behaviors or by what they say and think, that might shed some light on the degree to which goals were achieved.’ And in the process of doing that, I will probably discover other things as well.

LP: Um hum.

BS: Does that answer your question?

LP Yes, that was wonderful. Thank you

BS: Ok

LP: Great. We’ve been reading a lot of evaluations and talking to people about evaluation in the field, and one thing that we’ve been kind of struck by is that it seems challenging to meet evaluation needs while remaining authentic to the free-choice nature of informal learning experiences. And we were wondering what your thoughts were about that issue?

BS: Well, I think what comes to mind when you raise that question is how I like to keep on the broad brush stroke approach because-, [BS laughs] I have seen evaluations written by other evaluators where they say, 'These were the communication goals of this exhibition: Visitors will learn this, visitors will learn that. Visitors will understand the difference between this and that. Visitors will see the connection between this, this and this.' Where they get very very specific in the cognitive goals and expect the evaluation to prove that visitors learned these things. I think that's wrong-headed. It's both wrong-headed in terms of the designers, the exhibit developers to think that that's what an exhibition should do, and I think it's presumptuous on the part of an evaluator to think that they can evaluate for all of those very specific cognitive outcomes.

Because here's the deal- okay we get into my philosophy- that's not what exhibitions are about. With learners, and I'm going to call them learners, because they potentially are, the visitors. They are coming in with such a wide variety of motivations. They're coming in with such a wide variety of prior knowledge. And they're coming in with such a limited time schedule, a limited time budget. With those three factors, you can't treat educational evaluation of exhibitions the same way that you would treat educational evaluation of a program or a course that lasted three months.

LP: Uh huh, ok.

BS: But I still see people doing it. I wouldn't do it but I see people doing it.

LP: Ok, so that is a challenge. What are some ways to overcome that? Or what are other approaches that you would recommend?

BS: The approach that I take is, 'What was your big idea? What was the overarching message? What was the overarching content that drove the intellectual brain of this exhibition and the design team?' And very often there wasn't one. Very often there wasn't one. And I think that leads to a lot of the problems that a lot of exhibitions have.

So to make this work, it helps if the design team, the planning group, and stakeholders, up front come up with some kind of well-defined conceptual- it's like a thesis statement for the content of, 'This exhibition is going to be about--what?' What's the subject, what's the action, and what's the consequence of this idea that you have? And then you formulate the exhibition around that big idea. Instead of doing it in a hierarchy, a linear knowledge hierarchy of importance of ideas, you have one big idea and everything else is a satellite around that because visitors use exhibitions in less time and in out of order. It will not make sense if the exhibition is designed in a hierarchy. But it will make sense to a visitor if there's a core idea and that everything in the

exhibition in some way supports that. That way, they can use the exhibition quickly and incompletely and still understand the basic notion of what's going on.

LP: OK.

BS: And the cued questionnaire is directly related. It's an instrument that's also on a broad brush thing that says, 'What was this exhibition about?' It asks the visitor, complete these probes: 'It was to make people--what? It was to show--what? And it was to make people--what?'

So there's both sides of the coin that say, 'I think the intent of the museum for this exhibition was to communicate--this.' I'm being much more explicit than a visitor would be but that's kind of the nature of the answer to the question. 'And I think that the intention of the museum was to have it impact me in this way.'

So if you can get the answers to those two questions- 'What was this exhibition about? It was to show—what, and it was to make people—what,' -then you know whether visitors have the slightest inkling of what the big idea was and if they had the slightest inkling of what the museum's intended impact was. So there's two sides to the coin. Both what are they showing and what are they hoping the impact will be? It's like an output and an outcome. A lot of evaluations are about the output. 'We intended to make an exhibition this big with these features. We did. Here it is. X number of square feet, x number of exhibits, lots of interactions, lots of videos. Good press. People came. We were successful.'

LP: Ok. So kind of getting at that core idea for what drove the creation of the exhibit and then kind of satellite areas around that for visitors and so that they can kind of quickly get an idea of it without using it comprehensively.

BS: Yeah. If there's 24 or 48 exhibit elements, and they use 25% of them, will they come away understanding what's going on?

LP: Ok. And then to get at that, to use these cued questionnaire with broad questions about, 'What was this exhibit about? What did it show? And what did it want to make people do?'

BS: Yeah. The language of those two questions is in *Paying Attention*. And then going along with that is the tracking and timing data. The unobtrusively gathered data on 'How long did they spend?' Out of a population of people, a randomly selected population of people that visited this exhibition, what was the average time, what was the average number of exhibits they looked at? Where are the warm spots? Where are the hot spots? Where are the cold spots? And how does that overlay with the information that you get from what they say about the exhibition?

LP: Ok.

BS: So the tracking and timing data is like a real reality check. If people tell you in an exit interview, 'I read everything.' Hahahaha. [BS laughs and LP joins her in laughing]. And then you look at your tracking and timing data, and they didn't stop at half of the exhibit elements that were there. How could you read everything if you didn't stop? The first thing that has to happen is that they have to stop and pay attention. That's the first thing that has to happen. No attention, no engagement, no learning, no affect. Except on, maybe, a broad thing. I went into this exhibit. I didn't find it very interesting. I went through really fast. I stopped at 10 things. I saw everything I wanted to see. I read everything I wanted to read.

LP: Ok. So then those two pieces kind of coming together?

BS: Yeah.

LP: Ok, ok great, thank you for that. The worth question was kind of philosophical and I have another kind of philosophical question for you.

BS: OK.

LP: Sometimes it's said that evaluation is both an art and a science. How would you interpret that statement?

BS: Hmm. [Pause.] Is evaluation an art? Hmm. [Pause.]

I certainly think that planning, designing and making exhibitions has a lot of art to it. But I'm not sure that evaluation is an art as much as it is-, as much as it benefits from the broad experience of the evaluator and matching that experience with the goals of the institution for the evaluation. And then accomplishing it in a systematic way. And reporting on it in a clear way.

So as far as evaluation being an art? Am I more artful in my evaluations than other people? I don't think it's art. I think it's communication. And I don't think communication is as much an art as it is a talent, a skill that can be learned. I think evaluation is definitely a skill that can be learned. But there's many different directions to go.

LP: Okay. And what influences those different directions?

BS: Who you learn from. Who you hear and who you identify with and who you find credible and who you believe, and you know, what their philosophy is. Yeah.

LP Ok, ok great. So it sounds like you would describe it more as the broad experience of an evaluator and matching that to the goals of the institution. And then being systematic about the approach. And more so than an art, it's really more about being able to communicate and having skills that you can learn. And then those different directions that can be applied to evaluation.

BS: Yeah. [Long Pause.] Yeah.

LP: Ok.

BS: And I guess if you're gonna take a direction and say, 'This is my direction and this is why I'm selecting this direction and these are the benefits that I see for going in this direction,' That in the process of doing that communication, laying out those ideas clearly, you don't dis somebody else's direction in a way that mischaracterizes them....

LP: Ok. And you'd mentioned there's different directions you can go and there's different approaches. You also talked about matching the institution to the experience of the evaluator. People with different approaches are going to come into it differently. So how important is that process of matching the institution to the evaluator in getting a useful evaluation?

BS: Hm. I always thought it would be interesting to select a peer reviewed good exhibit. Ok, you'd start with an exhibition that is generally thought to be worthy and valuable. [BS laughs] Worthy and have merit. And then you invite two very different evaluators to evaluate it- that come from those completely different camps. So you have Evaluator A and Evaluator Z. And they're both given the challenge to do a summative evaluation of this using whatever methods you want. And then you just look at the results of those two studies and you say, 'How are they alike and how are they different?'

LP: That would be really interesting.

BS: Yes it would! But it needs to be funded. And it needs to be funded seriously well. And nobody's ever proposed that idea and nobody's ever come up with a budget for it.

LP: Ok.

BS: But this would be evaluation research of exhibitions. NSF used to fund summative evaluations and now they've sort of moved away from that and we're still floundering around here.

LP: Ok, wow, that's really interesting. I like that idea of same exhibit and different approaches and comparing them. It's really interesting. Ok great, thank you.

Speaking of comparisons, you know as we've been looking at some of the evaluation reports on informal.science.org and our general impression is that many studies are limited in their use of comparisons, such as either comparisons with other groups, or over time, or to other similar interventions. And we were wondering if you agree or disagree with that impression?

BS: Say that again please?

LP: Our impression, as we've been reading reports on informal.science.org, is that studies are limited in their use of comparisons. So either comparing one group to another group or one time period to another time period, like pre- post or to similar interventions in other organizations.

BS: So you're saying the results of the studies are hard to compare?

LP: No, just within the study, there aren't comparisons.

BS: Comparisons of what?

LP: Of that exhibit to anything else. Sometimes it seems like the evaluation is just this bubble of, '90% of the people liked this exhibit' but it's not compared to anything else.

BS: It's not compared to an exhibit where only 50% of the people liked it?

LP: Right. So is that typical of exhibits or is it not typical? How do we know what that means if we can't compare it to something else?

BS: Right. Ok. This is where we get back to the tracking and timing and the cued questionnaire. It's very very easy and solid to compare tracking and timing data across exhibitions. You can do that, but it's very difficult to compare the results of an open-ended questionnaire between exhibits because if the cued questionnaire instrument is kind of based on the assumption that the exhibition had a big idea and is fairly effective, right there you've got two really weak pillars that you're standing on. If your visitors were able to say with great relevance and emotion, that the exhibit was to show and was to make people-, and that what they say resonates positively with the intent, that's probably happening because there was a clear big idea, and it was really interesting and fun. And people got a lot out of it. And they got out of it what the museum expected them to get out of it.

But so many museums and so many exhibitions are so vague about their intent, and are so sloppy in gathering the feedback. And so, general rule: if one visitor got one idea, that means that it

worked. So we have a real problem with the evidence matching up with the intent in a systematic way. [BS laughs.]

So it's hard to compare the cued questionnaire answers to the first two questions of 'it was to show, it was to make people,' and say what % of people had a high resonance with the intentions of the exhibition when you don't have clear intentions. So you'd need to start with an exhibition that had clear intentions and compare it to another exhibition that also had clear intentions to be able to make any kind of a comparison. And it's just really shaky because there's such a wide variety of ways that people make exhibitions.

When I was doing *Paying Attention* and I was tracking and timing, I could track and time the data. I could say 'this many steps and this many minutes' or something to that effect, but it was often very hard to go back and look at the exhibition itself and say, 'Was this more about science or more about history? Was this highly interactive or not interactive?' The boundaries and the boxes that we try and put exhibitions into don't have clear edges.

LP: ... we were talking a little bit right now about comparisons and in *Paying Attention*--your meta-analysis of visitor time in museums, you were able to generate new metrics for analyzing tracking and timing data that have since been used for comparing different aspects of visitor behavior and museum use. So, first question around that was, what was the original impetus for that study?

BS: Let me see if I can think back. I think it was exactly your point about wanting to be able to compare something against something. So if they spent this amount of time, was that a long time? Or was that a short time? Or was that kind of a normal time given a sample of a large number of exhibitions?

So I'm not saying that the sample that I got out of a 100 initially and another 50 in the second round, I'm not saying that it's representative of the world. I'm just saying it's the sample I got and this is what I found. And I was amazed at the stability of the conclusions between the 1st bunch and the 2nd bunch. I did not have my eyes suddenly, you know, I didn't have a change of view from the conclusions of the 1st group and the 2nd group. The second group is the 'Paying More Attention to Paying Attention' online.

LP: We saw that, that was great.

BS: When I did *Darkened Waters*, I thought that it was a good exhibit. And I thought that there was really strong evidence that people spent a lot of time and used the exhibition intently. They were very engaged. Okay, I thought they were very engaged. When I published the results of the data from tracking and timing in the report to the museum and to the exhibit developers- the

museum and the exhibit developers were in different museums, Pratt was the museum in Alaska that had originated the story and the ideas and I was evaluating *Darkened Waters* when it was on view in Oakland, CA-...

LP: Oh, ok.

BS: ...when I showed and discussed the tracking and timing data, the reaction of the people who were stakeholders...said, 'Is that all??' They were disappointed in the amount of time that people spent and the number of people that looked at different exhibits. They, in their mind, had the impression, that people spent more time and did more stuff.

LP: Ok.

BS: So this meant that I had to say, 'Well, gee, [BS laughs] I think that in *Darkened Waters* people spent a lot more time and looked at a lot more stuff than other exhibits.' So now I had to go out there and say, 'Ok, who's got some other tracking and timing data, and can we make some sense of it and say whether or not this is unusually more or less or unusually less?' So it was a direct experience of people saying, 'Is that all?'

LP: Ok, that is really fascinating.

BS: Right. And it turns out in the big picture of the data that *Darkened Waters* is among the exceptionally thoroughly used exhibitions. But it's not the only one, we found some others. But when you look at those others as a group of 15 or 20, or however many, however few there are, it's very hard to see any one big silver bullet that comes out of that that says, 'Oh, it was because of this, or oh, it was because of that.' It's because of a whole range of different things.

LP: Ok, great, yeah, we love that matrix too that you have, the quadrants.

BS: Yeah, oh let me throw in another thing and that is that I had some discussions with people who were doing evaluations for the Smithsonian. The Smithsonian, this was my imagined idea, the Smithsonian saw that their sweep rate was off the chart in terms of other museums, so they didn't think that sweep rate applied to them. But I said, 'no no no no. You have your own data set. You should compare yourself to yourself. It's your museum, it's your visitors. Start collecting tracking and timing data. Start figuring out what your sweep rate is and compare yourself to yourself.'

LP: Yeah.

BS: And they have started doing that. They started getting excited about having an exhibition that has a slower sweep rate, especially when it was intended.

LP: Oh!

BS: If they compare themselves to everybody else on the quadrants, they look terrible. Because they're so big and people move through so quickly. So when they redid the mammal hall at the Smithsonian Natural History Museum, they made really wide aisles and they made really big vistas and sight lines. They said, 'We want our throughput to be high.' And then in another newer exhibit, they said, 'We really hope that people will move more slowly in this one.' And then in another new exhibit they said, 'Well, we did this and we tracked and timed in here, and oh look, we were more successful at slowing people down in this gallery.'

LP: Hum, ok, that's interesting, so more intentional?

BS: Yeah, if their goal was to get people to spend more time, and look at more stuff, be more engaged, they could use their own numbers, they could use the same metric, but they could use their own measures to compare themselves to.

LP: Ok, ok.

BS: But on that quadrant? I think that quadrant has some validity too. Even though people say, 'Well, it's not my museum, those aren't my visitors.' Well, unless you're the Smithsonian, you're probably more like others than you're different. But the Smithsonian is truly different because it's so big and people go there with 1 hour to see 5 museums. They don't do that in most cities.

LP: Ok, great, that actually leads nicely into our next question: what is your ideal vision for the use of those metrics?

BS: I think that they give you a really good tool for planning an exhibition, to be very specific and concrete about your expectations and your hopes for what visitors would need to do in order to get the idea or for them to feel successful. And then they give you that check when you're done. Did they? How close did you come to your intended thing?

I think that we waste a lot of real estate and we waste a lot of time making exhibits that are too big and too complicated, especially temporary exhibits. There's a philosophy that just says, 'We want people to feel like they got their money's worth, so we have to put in as much stuff as possible.'

LP: Right.

BS: So that's a different way of looking at it. I think that they serve as very useful metrics for comparing your own data against your own data or your own data against the world at large.

I also think they can be used as a planning tool. Just like the framework for judging excellence, is written in the past tense-- it's written to say, 'Did this happen or not? To what degree did this happen in this exhibition that's up and running?'-- But you can also change the tense of all the criteria, of all the aspects, to a future tense to say, 'The exhibition will do this and will provide this and will afford this' and use it as a planning tool. The cued questionnaire too can be used as a planning tool. You look at the cued questionnaire and say, 'What would I hope, what would give me the most goosebumps to read on this evaluation form after it's done? What are the answers that would make me just want to hug and kiss my visitors?'

And I think goosebumps and hugging and kissing visitors should be used as metrics. [LP and AF laugh.] There's a lot of museums that aren't even thinking about their visitors as being worthy of being hugged. They're just this sort of mass of people that they need to sort of tolerate, and wouldn't the museum be a lot cleaner and nicer and quieter if they didn't have them? [LP laughs.]

LP: That's interesting, I like the idea of a goosebumps metric. And that is our next question, and I don't know how serious you were about the goosebump-o-meter, but we were curious what other metrics or meta analyses are needed in the field? Do we need a hug meter?

[BS, LP and AF laugh]

BS: I think hugs and goosebumps are definitely something that we should look into. [LP & BS laugh] Because when I'm reading the raw data from the cued questionnaire and I read, 'It was meant to-, they wanted to show this, and they wanted to make people do that.' And if those answers just give me goosebumps because they're so right on and they're so wonderful, that has to be converted in some way to be a metric. I haven't come up with anything better than goosebumps right yet.

LP: That's interesting. So who would you be measuring the goosebumps on? The visitor or the person reading the feedback?

BS: Oh, it's the people that are reading the evaluation. [LP laughs] It's not whether the visitors said they got goosebumps. It's whether the reviewers of the data get goosebumps from what people said.

LP: I got you, ok, ok. Are there other things that you see? I think the tracking and timing is so interesting in so many ways. And I wonder if there are other things that can be done for comparison? Other metrics or analyses that would lead to metrics?

BS: Well, [pause] it's really really complicated. It's just so complex to look at the milieu of intention, capability, degree of execution.

I think it would be helpful to the field to be able to hear criticism more. There used to be sessions on the AAM called The Exhibit Critique. And a panel of professionals were selected and were asked to go into an exhibition in the town, in the city where AAM was being held, to review that exhibition themselves with their own opinions, their own camera, their own standards. And then come back and talk about the exhibition. I found it to be the most stimulating, exciting session and so did a lot of other people. It used to be packed. But there were complaints, there were complaints that the reviewers were mean [emphasizes word 'mean'] and that the creators of the exhibition should have an equal voice on the panel.

When I developed the framework for *Judging Excellence*, it is specifically in the absence of any direct help from the creators of the exhibit other than what they've put into that exhibition. In other words, you don't talk to the designer, you don't talk to the educator about what they were trying to do, or what went wrong and why they couldn't do something. Or why something is less effective than they intended it to be. You just look at what is there.

But I don't know how to get the framework into a wider platform. I don't know whether we could do an Excellent Judges discussion at AAM because everybody would want to play, everybody would want to have their opinion of course. But the point of the framework for judging excellence is that there are some widely agreed upon criteria. And that's what the criteria and the aspects are in the framework. These are things that people are not arguing about. These are things that pretty much are agreed upon.

However, having said that, the degree to which those aspects are seen and actualized in an exhibition depends a lot on the viewer's eyes and where they're coming from and how they're looking at the exhibition. So you can have a whole spectrum of agreement to disagreement among the judges, who are all using the same criteria, who are all using the same framework, but you can have differences of opinion of the degree to which it was accomplished. And I think that makes for a really really different kind of discussion and a really really different mood professionally to the discussion, than just a free for all review / critique. Where every individual person is coming at it with their own set of standards, unwritten, unrevealed, and their own professional way of looking at it.

In a good Excellent Judges discussion, people aren't saying, 'Well, because I'm a designer, I really think that this should be done.' They're not talking about it from an individual standpoint like that. They're saying, 'As a designer, I've had experience with this, and my reaction to this aspect and how it was realized in the exhibition is this, because I'm thinking in this way.' So the whole room is agreeing on the aspect and not disagreeing on the aspect, but it's your opinion about the degree to which it was accomplished that you saw differently. And that makes you really listen to somebody that's not agreeing with you in a different way.

LP: Ok. So if I'm hearing you correctly, it's more structure for that criticism?

BS: It gives structure to the discussion, exactly.

LP: Ok, and then just that need for more criticism and structured criticism to...

BS: They sort of tried to do that on the website *ExhibitFiles*. Are you familiar with that?

LP: Uh-huh. Yeah.

BS: Ok, so you can go on to *ExhibitFiles* and read my review of the exhibit Y-O-U, at the Museum of Science and Industry here in Chicago. I went with my friend Nancy and we go and we just tear them up. We're not using the Excellent Judges framework, we're just ourselves. And we really hate being made to feel stupid and we really hate being bored. So we attack. If we're being made to feel stupid, we attack. So we wrote, I wrote the review of You. And then it's an open forum that people can write back. Well, the exhibit developer wrote back and said that I was wrong.

LP: Uh-huh.

BS: [BS laughs and LP joins] Because I didn't understand what their intention was. So to say that your visitor is wrong? That doesn't make any sense. Excuse me? I'm sorry.

LP: [laughs] Ok, great, so forums out there for that kind of discussion?

BS: Need more, need more forums, yeah.

LP: Ok, great. I'm going to switch gears just a little bit here. I'm looking at the time. There are a couple more questions I want to try to squeeze in our last 10 minutes. And one is specifically about summative evaluations and funders.

Often funders ask programs to carry out summative evaluations; and of course, the programs comply. What do you think the funders are looking for in those summative evaluations?

BS: Well, I'll tell you what I wish they were looking for. I wish that they could feel good about how much money they spent and that they would want to spend it again. And that they would be respectful of the challenges that museums face going into this and that they recognize that we don't know what we're doing, really. We have some pretty good ideas. But when Roger Miles at the natural history museum in London, many many years ago, this was like 30 years ago, first decided to start incorporating evaluation in the planning and design of exhibitions and summative evaluation, he received such criticism and pushback from people who said, 'I thought you knew what you were doing? I didn't think you needed evaluation, because I thought you knew what you were doing. But you're now admitting that you don't know what you're doing? That you need to ask visitors to find out whether you're doing a good job or not?' Well that was radical then. It's still radical now with a lot of people. And I think maybe funders are kind of the last to know about this, that you really do need to talk to your visitors about it, and you really do need to get that feedback. And you need to get it at a time when it's useful to make changes, to make it better.

So not spending any money to do any front-end evaluation or any formative evaluation, and only spending a big wad of money at the end to do a summative evaluation, and not saving a percentage of the development costs out to make remedial changes after you've done your summative evaluation to improve things, it's not money well spent if you only spend money on summative evaluation. So I think funders need to understand that too.

LP: So in an ideal world, what do you think summative evaluation should be? How would that look to you in an ideal world?

BS: I think in an ideal world- Hold on a second, I'm digging up something. I can send you this if you want. Or you can probably transcribe it from what I say.

I believe that in the best possible world, this would be what summative evaluations would be about. And I'm reading this as a first paragraph of my abstract for the BISE report. Ok. This is taken largely from the informalscience.org CAISE website. I think they say the same thing.

“Summative evaluations of museum exhibitions are generally conducted with the aims of measuring whether an exhibition met its goals, identifying areas for improvement, and assessing impact. In many cases, evaluation studies also serve to advance the field by deriving lessons for funders, policymakers, or practitioners beyond the project.”

I think if every summative evaluation did those things it would be lovely.

LP: OK great. Thank you for that. And then a follow-up question from that: If that happened, then what is the best-case scenario for how those summatives would be used by stakeholders?

BS: Their own? The ones that they've funded?

LP: Yeah, I guess whichever stakeholders you want to consider, whether it's the funder or the institution or...

BS: Ok, you like lots of stakeholders.

LP: Yeah.

BS: Well, I think they should read it and they should be able to get all those things from it. They should be able to know the degree to which the exhibition met its goals. First of all, they should know what the goals were and whether those were reasonable or not. They should know what areas were perhaps missed opportunities, where there could be improvement, and they should know what the impact was on the visitors. They should have some real good goosebump producing data that witnessed something good happening.

And it would be nice if there could be lessons for practitioners beyond the project, if there were some lessons that were easily shared.

LP: Ok, and it is so unfair to ask this with so little time, but what about the role of recommendations?

BS: [laughs]

LP: [laughs] I know, I'm like, oh no, this is so unfair with such little time because it's the focus of your...

BS: Do it! Do it!

LP: So what is the role of the evaluator in providing recommendations in those ideal summative evaluations?

BS: I think that's where you identify the areas for improvement, I think that's where you derive lessons for other people, derive lessons for both the people that did the exhibition and for people beyond the project.

LP: And just to get you on the record, do you believe that the evaluator has a responsibility to do that as part of their summative?

BS: I think it makes for a much more interesting and useful report. I know some evaluators that say, 'Oh no, no, no, I can't do that, I don't feel that's my role. I think that the client should decide what the recommendations are.'

I think that, speaking personally, I have a lot of experience, and my clients want to know what I think. Maybe some people don't have a lot of experience and don't feel comfortable saying what they think. So I think it has to be probably an agreement. I know, I have, in the past had it pointed out to me by clients, 'We want to know when you're talking about what the data say and when you're talking about what you think about that.' So they do want me to be clear in saying, 'This is what the data say and this is what I think the implications of the data are.'

...

LP: Thank you for your time in this interview, it's been very valuable for us.

[END INTERVIEW]