

Discussing the Framework for Evaluating Impacts of Informal Science Education Projects

Multiple Deliverables

Participants:

Charlie Walter
Luisa Rosado
Saul Rockman
Cynthia Needham
Paul Dusenberry
Bob Hirshon
Val Kass
Sue Ellen McCann
Robert Blair
Carol Lynn Alpert

Sue Ellen McCann (*Quest* Program) — I checked all the boxes in the survey. This was alarming because our project wasn't structured for in depth evaluation on all those aspects. Should we have a particular focus or include everything? We have TV, radio, web, formal and informal ed., community events, partnerships, and kiosks.

Question: If you have a project with multiples, is the on line software configured for all these different goals and measures??

Robert Blair: I was one of the guinea pigs (*Master Naturalists* program). It was fairly doable. Not painful; in fact, it was quite nice. And we had multiple components, curriculum material, training materials for the volunteer. So we have multiple audiences: trainers, trainees, lesson plans, visitors to museum.

Bob Hirshon: I think we are dealing with two different sorts of projects: some have a variety of components that all work together to achieve one set of coherent goals. In that case, you can just evaluate whether or not the suite of components, together, did what they were intended to do. But in other projects, you have a variety of components, each of which is designed to do something different. For example, you might have a TV show that reaches a public audience, and an additional component is using elements from the show as the core of an after school program. In that case, you have different audiences and different outcomes for the two elements. The evaluation needs of these sorts of projects are completely different.

Val: It also comes down to audience. Do they have the same audience or different audiences? With Cyberchase, there is a project just to try to measure how the components work synergistically or on their own.

Saul: Also, there is a difference between the outcome of a 20- minute museum exhibit vs. a series of events over a period of time. So the dosage of the different elements is also important.

Carol Lynn Alpert: What about simultaneously submitting to ISE and also to the DRE research group to conduct research on our projects? Are we considered ISE researchers? Or are we ISE practitioners who hire others to do research?

Val: The Cyberchase research grant was submitted by an evaluator, not the producers. But this is something we should discuss.

Paul Dusenberry: You sometimes have one or two dominant deliverables, and then other minor ones.

Carol: Right, will it be acceptable to say we'll just evaluate our main deliverables?

Margaret Burke: It's also important to evaluate the process, in addition to the deliverables.

Saul: Yes, how the different teams for these deliverables work together is important to look at.

Val: What the evaluation should cover depends on the teaching goals you have for your project. The evaluation has to match your goals. (talks about a case where they negotiated with a producer, because their goals and evaluations didn't match)

Carol: So there may be more opportunity for negotiation and dialog on this topic between NSF and the producer to go back and readjust?

Val: We always do that; yes.

Bob: There is also the matter of what's already been done and what's new. The main deliverable of our project may be something that has already been evaluated to death, but one or more ancillary deliverables—minor things that rely on the main deliverable, and are spun off from it—may be new and interesting and have less of a research base. So then the question becomes, should we focus on the main item, which covers many of our important outcomes, or do we focus on the cool new stuff?

Val: It's not one size fits all. It often turns out the continuing projects sometimes focus on the new things.

Saul: You want to start with your impact, and then how you'll accomplish it.

Val: What you want to look at is what impact do you have on the population you target.

Saul: (regarding *Quest*) Sampling makes a difference. How is the audience differentiated?

Sue Ellen: Our project will result in a community baseline. No one else has been gathering this information. In the future, we may be able to use that baseline to draw better conclusions about the effects on different facets of this audience.

Margaret Burke: It goes back to accountability. With these broad evaluations of many topics, it's hard to tease out what is having the effect.

Sue Ellen and Saul and Meg: The biggest effect of project with multiple deliverables could be the breaking down of these silos and creating valuable collaborations—those who work on the different deliverables, some who are from different institutions, are working together in new ways. But this is hard to document.

Sue Ellen: In addition, beyond the various deliverables and outcomes we start out with, the project grows and evolves

Cynthia: But the system seems like it will accommodate these; it's pretty flexible. And we can add objectives later

Saul: The framework allows a lot of flexibility. But your goals have to be realistic. We often see proposals with unattainable goals, like “children who participate in this project will grow up to be microbiologists.”

Carol: Maybe our impacts happen much later, after the grant is over. Many scientists say “I got into science because of my trips to a science museum.” But we can't capture those effects. What can be done?

Val: Sandy Welch and I were just discussing long term outcomes: Wouldn't it be great to do a long term study? For example, the NOVA show on Evolution done years ago still shows up in searches near number one. It's still having an effect, ten years after it was produced.

Carol: But that's more like shelf life: I'm talking about effects that don't show up until later.

Meg Burke: There was an evaluation of NSF ISE programs: They asked practicing scientists what inspired them. 80 to 90% pointed to informal science experiences. Maybe we could do an evaluation like that again?

Saul: There are immediate outcomes; then there are long term ones, 3 weeks later, what do you remember; but it's often the case that once the PI is done, she/he wants to just move on. They don't want to save some of their evaluation money for an extra year of evaluation. They want to close the books.

Meg: How about supplements? To take care of unforeseen impacts, or to extend it for research purposes?

Paul: The supplement rules are pretty strict...

Val: They are case by case. It really depends on funding.

Meg: Or, on the other hand, there are planning grants. Maybe we'll all need to do planning grants to cover the cost of developing really extensive evaluation plans.

Saul: The planning of evaluations is a lot of work. We just did one 9 pages, another one 7.5 pages.

Robert Blair: And it's a problem that the reviewers don't even see the evaluation in the appendix. All that work is done, and it gets summarized in a half page in the proposal.

Paul: I've heard that evaluation firms will become gatekeepers. Everyone has to find an evaluator. If they pitch them and they don't want to take it on, that can stop it right there.

Hilarity ensues...

Everyone: So, Saul, how much will it cost me under the table for you to take my project on?

Paul: The other problem is the independence problem: Should the summative evaluator be different than the formative?

Saul: It depends; it's not necessary in all cases.

Sue Ellen: The idea of coming up with a set of best practices, even if it's just a starting place, would be valuable.

Louisa: The framework is a useful guide to tell us how to go about it but also what we'll be held accountable to. But with multiple deliverables, perhaps you have to just choose one to evaluate and not worry about the others.

Saul: You can evaluate them together; you can begin to lay out what each will do, and how they work as additive elements. We have evaluated elements together and as stand alone to see what additive effects there might be. And we've found additive elements don't really add anything.

Sue Ellen: It would be good to have suggestions as to how to go about it; what questions to ask; examples of similar projects and what they found.

Meg: If we got standard questions for each bucket, or at least standard headings, then we could combine our results, and create get large- scale datasets we can contribute to. How

about a brainstorming section with experts? What are the standard questions we should ask about, say, attitude. This would standardize the information.

Saul: I still say “it depends.” But if you pull together a bunch of firms, maybe we can agree on a certain set of questions. That may result in something usable.

Val: What would it take to do that?

Saul: I’ll talk with Irene (Goodman) and we’ll put together a proposal.

Val: It is uncertain that aggregating it all will be helpful: it could be apples and oranges and grapefruit, because the projects are so different.

Robert Blair: How about 20 perfect questions for each bucket, as examples.

Paul Dusenberry: Many science organizations get together and do common data formats. So we could have at least common standards of data.

Charlie Walters: There are very few standards in exhibition projects; some are process, some are content. Also, we’re looking for innovation, and this sort of evaluation forces everything into the same mold.

Bob: At Project 2061, they are taking all of their learning goals—hundreds of them—and working out valid assessment questions for every one. And they’re compiling questions that DON’T work, too, and explaining why they either lead a kid who knows the material to answer wrong, or a kid who doesn’t know the material to answer right, etc. So they are systematically looking at knowledge assessment. It’s possible that that work could inform our work on valid questions for our outcome measurement.

Two Key Points from the discussion:

With different deliverables and different audiences, is there a way of articulating the synergy? How all the pieces work together?

This will be an opportunity to look at the process of building collaborations, since different groups are working together.