SCHOODIC INSTITUTE AT ACADIA NATIONAL PARK

Research Report

Pathway to BioTrails: DNA-Assisted Species Identification for Citizen Science

What the Project Learned About How DNA-assisted Species Identification Might Add Value to Learning

Bill Zoellick
Schoodic Institute at Acadia National Park
September 30, 2015

Executive Summary

The Pathway to Biotrails project explored whether and how DNA-based identification of organisms, known as 'DNA barcoding', can extend and enhance informal science learning in citizen science projects involving adult volunteers.

It began this exploration by investigating a carefully specified model for engaging adult volunteers in citizen science fieldwork and learning. But as work proceeded, we became increasingly confident that DNA barcoding's impact on citizen science learning would not just be limited to one particular kind of event or approach to organizing volunteers' participation and learning. We recognized that future applications of DNA barcoding would support learning in many kinds of events extending over a variety of durations. Consequently, the second year of the program turned from a focus on one particular way of organizing learning to an exploration of a variety of models for participant engagement that might be supported by DNA barcoding.

In addition to studying participant learning in the initial event design, which extended over two successive weekends, we studied participants in a one-and-a-half day specimen collection and sorting event, in an hour-long classroom event, in the context of a bioblitz where there was ongoing sorting of specimens, and in a short field collection event organized by the National Park Service.

In order to characterize patterns of participant engagement across these different approaches to structuring participation and learning, we developed an observation protocol that could be used even in very short duration events where there was minimal opportunity for interviews and other direct interaction with visitors. Even though the protocol still needs additional development and refinement, we found that it was useful in sharpening attention to differences between events. It is an outcome from this work that may be useful in subsequent research.

Background technology or in the foreground for learning?

This exploratory study provided evidence in support of a number of assertions that will be helpful in informing future learning research associated with DNA barcoding. Perhaps the most important assertion is that DNA barcoding provides "background" support for learning in that it will, over the next few years, enable more citizen science research that is dependent on species identification. We also explored the question of whether DNA barcoding can usefully support learning in the "foreground," where the learners either engage directly in DNA barcoding or learn about it as a topic of interest. We found that DNA barcoding was interesting to many participants and was something that they said they enjoyed learning about. But we found that it is more something that citizen scientists "learn about," as opposed to "learn to do."

The important implication of DNA barcoding's support in the background rather than the foreground of citizen science learning is that the important questions regarding learning design and outcomes will be related to the particulars of the citizen science work that DNA barcoding enables, rather than to DNA barcoding itself. Put another way, there is no single, inevitable or necessary connection between DNA barcoding and learning. Instead, DNA barcoding will have the effect of enabling many new and different citizen science investigations that focus on biodiversity. The quality and amount of learning associated with those investigations might be good or bad, large or small. The learning will not depend on the DNA barcoding. It will depend on the learning design of each different project.

Support for learning though more citizen science biodiversity research

This is NOT the same as saying that the rapid maturation of DNA barcoding technologies has no impact on learning. Increased availability and decreased cost of DNA barcoding *will* impact learning because it will enable more biodiversity research. The important question, then, is how to support learning in biodiversity events.

Event duration and research on learning

This exploratory project also helped us develop an understanding of the advantages and disadvantages associated with different durations of events. We did not find a relationship between event duration and number of participants in the events that we studied. But we did find that as events became very short, it was increasingly difficult to draw conclusions about the nature of the participants' engagement in the event or about what they learned. We found that in order to support learning research, an event length of a day or two was long enough to support the collection of rich data. Another workable approach would involve a group of participants in a series of shorter events repeatedly, over time.

Interactions between experts and participants: a research opportunity

This study also offered potentially useful insights into the interactions between experts and participants. The scientists, subject matter experts, and others leading the events that we observed typically answered questions authoritatively, or chose not to answer them. With one exception, they did not engage participants in the kind of guided inquiry associated with good classroom teaching. Their role as "experts" also precluded mutual learning among scientists and citizens as peers. This is not a criticism of the role that scientists and other experts chose to play in these events. The events were of relatively short duration and, in some cases, focused on accomplishing as much sorting and other work as possible within a limited time. Further, participants expressed interest, along with the scientists, in getting as much work done as possible. Clear leadership makes sense in such settings. The point here is that this question of the nature of the interactions between volunteers and scientists and experts appears to be a potentially rich focus of future inquiry into learning in these kinds of events.

Recommendations

This report concludes with a number of recommendations that might be useful in structuring future research into understanding and improving the learning that happens in citizen science research enabled by DNA barcoding. Briefly summarized, these recommendations are to:

- Deveiop clearly stated conjectures about the learning that will be supported by the citizen science activity.
- Avoid broad outcomes such as changes in "attitude" or increased "motivation" that are empirically ambiguous.
- Articulate a detailed "theory of action" at the outset. The theory of action should consist of a set of testable conjectures about how the citizen science activity will produce the intended learning. The goal is not to test whether the theory is "right." Instead, as conjectures are tested and modified in the face of empirical (not necessarily quantitative) evidence, the theory of action changes too. It is a guidance mechanism for the research.
- Identify "black boxes" and look inside them. Projects that include a significant implementation component, in addition to research, often focus on doing something and then seeing

what happens. What is missing is an understanding of why. Attention to black boxes as they pop up in the course of implementation helps avoid this.

- Place each study in the context of a broader suite of studies that are designed to learn from each other. The questions that need to be explored here are too complex to address in any single study.

Table of Contents

Figures

Introduction: Exploring a New Technology

In reflecting on what Pathway to Biotrails ("Biotrails") learned about informal science learning, it is clear in hindsight that the project evolved into an exploration of how the important new technology of DNA-assisted species identification ("DNA barcoding") might add value to learning in a variety of models for citizen science participant engagement.

This was not the project's initial design. But it seems to me that this "evolved" design was particularly appropriate for an exploratory, Pathways project focused on increasing our understanding of how a groundbreaking new technology might create new opportunities for citizen science learning.

This report describes what we did and what we learned in the course of this exploration. It concludes with reflections and recommendations that might be useful in shaping future research.

The Initial Design

The first paragraph of the Biotrails project proposal promised to, "explore whether and how DNA-based identification of organisms, 'DNA barcoding', can extend and enhance informal science learning in citizen science projects involving adult volunteers." The project did this.

Where this exploration evolved beyond the initial design was in the details of just what the adult volunteers would be doing. The initial design would have brought groups of adults together for four different five-day investigations, each spread across two weekends. These adult volunteers were to be divided into three different experimental groups. The first group would engage in the kinds of activities that are common in many citizen science species collection and identification events ("bioblitzes"): they gather samples and try to identify them by morphology, but do not get involved in analyzing and interpreting the results. The second group would, like the first, collect samples and try to identify them morphologically, but would also engage in analyzing the findings and thinking about what they mean. The third group would NOT engage in morphological identification and would instead spend their time extracting DNA from samples and preparing the DNA for sequencing. This group, like the second, would have the opportunity to think analytically about the results.

In short, the initial design had the structure of a controlled experiment with three "treatments." The design explored the conjecture that if DNA barcoding could replace morphological identification, it would open new learning possibilities. On the other hand, we recognized that examining organisms under a microscope and trying to identify them morphologically was, itself, a learning activity. Dividing the volunteers into three groups was intended to enable us to explore the different kinds and degrees of learning afforded by the different experiences.

Implementing the Initial Design: What We Learned

Recruiting Participants

The first thing that we learned was that it was difficult to recruit new volunteer participants who were not already involved in citizen science activities at MDI Biological Laboratory or the Schoodic Institute. Each institution has ongoing programs that engage local volunteers in various sampling and restoration activities. These volunteers could have comprised a ready pool of participants for this study. But these volunteers had already received training of various kinds. This previous experience and learning would, of course, affect the kind and degree of learning that would happen in their Biotrails experience. What is perhaps even more important in terms of our initial focus, these volunteers would not be representative of the millions of people who visit Acadia National Park each year. We were interested in investigating the kind of learning that might be experienced by park visitors and others who are not already connected to our organizations.

In hindsight, the difficulty in recruiting volunteers is not surprising. We were asking people to commit two weekends to this work. For people travelling from other places, this would mean arranging for other activities in the intervening week. There are plenty of things to do in a week at Acadia, but making such arrangements takes time and advance planning. What we were envisioning involved having participants organize the kind of vacation that would need to be planned months or perhaps even a year in advance.

Over the past year the Schoodic Institute has developed a partnership with Earthwatch, an organization with a long history of organizing expeditions that bring volunteers from around the world to work together with scientists and to assist with research. Earthwatch demonstrates that such collaborations between scientists and people who are on vacation can happen. But Earthwatch also illustrates the amount of promotion, infrastructure, and investment that is required to make such events happen.

Ultimately, we were able to recruit 13 volunteers to participate in a two-weekend event that implemented some of the elements of the initial design. We relied on personal relationships and on a mailing by Friends of Acadia, a local organization that provides support to Acadia National Park, to recruit these volunteers. Consequently, although the volunteers were not people who had participated previously in similar projects with MDI Biological Laboratory or the Schoodic Institute, most did come from the communities on or close to Mount Desert Island.

Learning Research vs. Field Ecology Research

Before the start of this first field event we decided to reduce the number of treatment groups from three to two. Two considerations informed this decision. First, and most importantly, we recognized that excluding one of the treatment groups from analysis and reasoning about the problem being studied, as initially planned for the first, “traditional bioblitz” group, would substantially diminish the experience for these volunteers. The focus of this fieldwork would be on collecting samples to assist in exploring the interesting, vexing question of causes and possible effects of rapid decreases in the amount of local eelgrass habitat. Just having volunteers collect and sort samples without engaging them in thinking with scientists about possible causes and effects seemed unfair.

A second consideration was that, with only thirteen participants, we would have only four participants in some of the treatment groups.

Once the fieldwork and subsequent sorting in the lab got underway, other issues began to emerge, as often happens in fieldwork. Some of the samples contained more mud than anticipated, which required more time for preparation and searching for samples. Some of the collectors had fewer specimens than anticipated. Some of the eelgrass blades required more counting of epiphytes than anticipated. The biologists leading the fieldwork huddled to figure out how to reorganize the work to deal with these changes. Some things needed to be speeded up. Other tasks were set aside or modified to accommodate the work that needed to be done within the time that was available.

One of the effects of these changes was that the clear separation between the treatment groups disappeared. Rather than having one group experiencing the event through interaction with taxonomists and morphology and another group whose experience did not include morphological analysis, all participants spent a substantial amount of time working with taxonomists and identification keys. Rather than being able to compare two groups of volunteers who experienced citizen science in clearly different ways, the difference was now just that one group would do morphology for a longer period of time while the other learned to extract DNA.

As a social scientist, rather than a biologist with a strong interest in what might be learned about the eelgrass, I found myself in the position of watching this rearrangement of what participants would do and what they would experience with interest. The interaction I was observing seemed to me to be an example of the tension between objectives related to the scientific work at the center of a citizen science project and objectives related to learning. This kind of trade-off and tension between scientific outcomes and learning outcomes has been documented elsewhere (Berkowitz, 1997; Penuel et al., 2006; Zoellick, Nelson, & Schauffler, 2012) and so is not surprising. In this instance, I recognized that achieving as much as possible in the way of outcomes related to eelgrass research was not only a primary concern for the biologists, but was also what the participants wanted to do. But I also recognized that it meant that, for this event, we would need to shift from an approach to learning research that deployed a carefully structured experimental design to one that depended more on naturalistic observation.

This is not a criticism of the program’s operation on this first weekend of fieldwork. I think that the decisions that the biologists made to rearrange the work plan were necessary. The important point is that this is an example of the kinds of things we learned in the course of this Pathways project that should be taken into account in future work. I see two lessons here:

1. In citizen science, when the value or integrity of the scientific work is at risk, learning concerns must take a back seat.

2. The design of learning research for citizen science should tend toward more flexible approaches that can accommodate shifting demands that arise on the non-social science part of the research. Social scientists undertaking research into citizen science are scientists studying another kind of science. To study it without changing it, unobtrusive research designs will often be the most useful designs.

Research Findings

This initial event was of relatively long duration, starting on a Friday and continuing until Sunday and then picking up again on the following weekend. This permitted collection of different kinds of data, including surveys and interviews in addition to observation. Appendix A of this report summarizes what we learned through interviews and observation. Appendix B analyzes the survey data collected at the start and end of the event.

Briefly summarized, the surveys showed that this group of volunteers came to the project with an unusually high degree of self-perceived comfort and knowledge about science. Several of the participants were science teachers; one was a retired, prominent anatomist.

Perhaps because these volunteers came into the project with relatively high levels of science knowledge, the surveys did not detect large changes in science knowledge or self-efficacy. There was, however, some evidence of improved self-perception regarding scientific inquiry skills that include matters such as "choosing a suitable site for gathering samples," "accurately identifying species for scientific study," "classifying or sorting specimens," and "using equipment for a scientific study (e.g., nets, microscopes, sieves, forceps, etc.)

We also gained important insights from observation and interviews:

- The participants were interested first in doing something useful and secondarily in taking home some new learning or experience. The biologist leading the eelgrass study set up a compelling problem and made the case that she needed help. To the extent that participants expressed frustration, it was primarily because they were not sure that their work was adequately contributing to new understanding and a potential solution.

- The participants in the DNA procedures group worked mostly individually after receiving instruction; the work in the morphology group tended to be more social, with people working in pairs and with the whole group moving around to see each other's specimens.

- Having participants engage in DNA-related procedures provided them with a different kind of learning experience (following a DNA process recipe rather than using a key to identify species), but it did not result in a more productive experience from the standpoint of serving scientific objectives. (Many of the gels that participants prepared failed.)

- Neither group addressed anything beyond species identification; they did not, themselves, have the opportunity to learn and think about ecology or systems of relationships in the eelgrass community. At the end, they did have the opportunity to listen to a conversation among the participating scientists about such matters, but the volunteers' participation was passive.

- The morphology group and the DNA extraction group saw their work as vaguely complementary, though the exact nature of the complementary relationship was not quite clear.

- Both groups were concerned about producing results that were truly useful to the scientists in the study.

- Both groups expressed satisfaction with their experiences.

Exploring Different Models of Participant Engagement

Our experiences in this initial workshop transformed our thinking about the project and about the course of the investigation we were pursuing. The experience sharpened our recognition that we were pursuing the specific question of what kinds of learning might be enhanced by the availability of DNA barcoding. This implied that we needed to look not only at the kind of learning or amount of learning, but also at the different learning contexts that DNA barcoding might enable. Our overall goal was to undertake exploratory work that could support design and implementation of ways to use DNA barcoding in citizen science learning. Our initial design focused on "going deep" in studying one particular model of engagement. We now realized that we should broaden our investigation to consider different models of engagement.

Further, our first workshop suggested that "long form" engagements might tend to be more attractive to people who are local and perhaps already have some experience with and interest in scientific work. Could we engage more people with less pre-existing interest in science if we used DNA barcoding as a way to couple useful scientific work with shorter-duration events? Quoting from the report included here as Attachment A:

> *After observing this past workshop we conjecture that there might be many different kinds of [citizen scientist] engagement. … For some (probably local) people, participation could be year after year and could involve thinking about the design of the study, the logistics of implementation, and the analysis of results in collaboration with the lead research scientist. At the other extreme, we might seek to involve visitors to Acadia in useful work that might be accomplished over a period of a*

> *few hours. This work might be in the field (e.g., supervised sample collection) or in the lab (e.g., sorting specimens from debris). We would seek to use DNA barcoding as tool to allow us to do truly useful work in many locations at many times while engaging participants that range from fully collaborative work [with scientists], on the one hand, to something more like a museum visit, on the other.*

Implications of Exploring Different Models

The shift from a focus on a single approach to structuring participants' interactions to a focus on exploring a variety of models for such interactions had implications for the structure of our learning research.

Perhaps the most obvious implication was that moving from five days of interaction with participants to interactions that might last only an hour or two meant that we would need to collect data in different ways. Taking fifteen or twenty minutes at the start of the session for participants to complete surveys and then repeating the survey at the end of the session does not make sense in events that last only an hour or two.

A second implication was that we needed to think in more general terms about what the participants might gain from their participation. Our initial plan to involve participants in one specific kind of event structure enabled us to think in terms of well-defined learning outcomes. Exploring participation across events with different structures and different kinds of activities pushed us to think more broadly about what participants might gain.

Our decision to think more generally and broadly is consistent with the exploratory nature of the Pathways project. But it changed the nature of the conclusions that we are able to draw from the work, moving them from specific conclusions about a particular kind of interaction to more general conjectures about ways to couple DNA barcoding and citizen science learning. The decision also necessitated development of some new ways to collect data.

Shifting the Focus from 'Learning' to 'Engagement'

Perhaps the most far-reaching implication of the change in our approach was that we shifted from thinking about "learning" to thinking about "engagement." Both of these terms are slippery and mean different things in different contexts and to different writers and readers. So, it is important that I clarify what I mean.

The proposal for the Biotrails project framed learning through use of the "Strands of Informal Science Learning" that are identified in *Learning Science in Informal Environments* ("*LSIE"*) (National Research Council, 2009). Specifically, it proposed to impact learning in the following strands and related behaviors and capabilities:

- **Strand 2:** Come to generate, understand, remember, and use concepts, explanations, arguments, models, and facts related to science.

- **Strand 5:** Participate in scientific activities and learning practices with others, using scientific language and tools.

- **Strand 6:** Think about themselves as science learners and develop an identity as someone who knows about, uses, and sometimes contributes to science.

We intended to assess whether this learning was happening through observation (for Strand 5) and through self-report by participants in response to survey questions (all strands). The surveys asked participants to assess their level of skill on a variety of science-related activities and their sense of self-efficacy in the area of science on a scale of 1-5. The surveys are included in Appendix B.

As noted above, the shift in focus to shorter-duration events made it impractical to administer surveys before and after events. We relied instead just on participant observation, supplemented when possible with conversations with participants during the program. We focused our observations on making judgments about participant engagement.

McCallie, et al. (2009), in a report by the Center for Advancement of Informal Science Education (CAISE) on public engagement with science and informal science education, address the ambiguity in understandings of the term “engagement:”

> *In terms of informal science education, engagement is a loosely defined term generally referring to behaviors that demonstrate interest in or interaction with a science-related activity or experience. Engagement is often considered an integral part of participation in or learning about science, or as a stepping stone to further participation or learning. However, how the term engagement is being used in a specific instance is rarely well defined. (p. 20)*

McCallie, et al. go on to develop a rich definition of public engagement that incorporates mutual learning by publics and scientists. I will return to this idea, reflecting on interactions between scientists and public participants later in this report. The definition of engagement that we used to structure our observations was less ambitious. It focused primarily on evidence of learning that is associated with Strand 1 as defined in *LSIE*:

- **Strand 1:** Experience excitement, interest, and motivation to learn about phenomena in the natural and physical world.

LSIE goes on to make the following observations about collecting evidence of Strand 1 activity:

> *In studies of informal learning, interest and related positive affect are also often inferred on the basis of behavior displayed. That is, participants who seem engaged in informal learning activities are presumed to be*

> *interested. In this sense, interest and positive affect are often not treated as outcomes, but rather as preconditions for engagement. (p. 61)*

In order to standardize our work as we observed participants and gathered evidence of "interest and related positive affect," I developed a simple observation protocol (Zoellick, et al., 2015) that we then applied and tested in a variety of informal science settings over the second year of the Biotrails project. A paper presented at the 2015 Citizen Science Conference in San Jose that describes this protocol and subsequent investigation of its properties is included as Appendix C.

In reviewing this progression from "learning" to "engagement," it is clear that as we expanded our exploration to include different settings and different approaches to interacting with volunteer participants, we reduced the number of *LSIE* learning strands that we attended to. The rationale for this reduction in the breadth of focus includes:

- We learned in our initial workshop that engaging participants directly in DNA extraction was not productive or efficient. Since part of the attraction of DNA barcoding is that it reduces the time required for species identification, this kind of participant engagement undermined one of the advantages of using DNA barcoding techniques. Because participants were so clearly focused on doing work that was actually useful, rather than just on opportunities to learn that did not contribute strongly to scientific objectives, there was no reason to adhere to the original research design.

- Doing away with the time required to engage volunteers in DNA extraction opened the possibility of focusing on events of much shorter duration, including those that focused primarily on sample collection with only minimal attention to species identification.

- For us, the primary goal in this project was NOT to produce rigorous research into learning. That would come later, once we had a clear model for participant engagement. At this point our goal was to gain experience with different ways to use DNA in support of citizen science so that we could think carefully about the design of a subsequent full-scale research project that WOULD result in rigorous, replicable results.

- We recognized that in order to explore different kinds of learning settings of different durations, we needed to develop some new tools. We turned our focus to that work.

Implementation and Research in the Second Year

A Shortened Field Experience

In mid-June, 2014 we again recruited volunteers to assist with collecting and sorting samples of eelgrass blades and organisms in an eelgrass bed. This second workshop differed from the first in a number of ways.

- It was much shorter. Participants gathered for a two-hour introduction and orientation from 3 to 5 PM on Monday, June 16, collected samples the following morning, sorted and preserved samples for two hours after lunch, and then assembled in a classroom for a brief introduction to DNA barcoding, a debrief, and wrap-up.

- It involved kayaking out to the collection site rather than walking to it, as in the first session.

- Participants did not engage in DNA extraction, just in two hours of using microscopes to capture and process organisms found on or around the eelgrass.

This workshop attracted six participants. Two of the participants came specifically because of the opportunity to do some kayaking. One said, “I have to admit I was here for the sea kayaking. [Friend's name] called up and said, ‘Hey it's just sea kayaking. We collect and ...’. I thought we were just going to collect for the Biolab and be used kind of like … as like … you know ... I expected five gallon buckets of sea water and we were just lifting and trudging all these. I didn't even know it was eelgrass.”

One of the participants was a local, retired engineer who regularly attends scientific and conservation programs in the area. He was accompanied by his wife and by a graduate student from Mexico who is involved in a master’s degree program in conservation management. The sixth participant was a psychology professor from the Bangor area.

Evidence of Learning: Strand 5

Despite its shorter duration, the second workshop succeeded in engaging all of the participants in learning about scientific activities and use of scientific practices and tools. A few quotations from a conversation with participants at the end of the session illustrate that seeing the world through laboratory-quality microscopes opened new learning opportunities for a number of the participants.

- *[I enjoyed learning] the technical stuff, just learning how to fiddle around with it to get the lights right and the focus right. And even using little tools, I got to scrape off something that was alive and get it into one of those little tubes, and it was like chasing a kitten you know because it kept moving around. So, it wasn't like some simple little things. But, that was like the practical aspects of doing it.*

- *It's interesting, the placements of the scalpel on the blade of grass and then looking in and then making sure you're not smashing what you're trying to … you can't really see with the naked eye. So, I mean it's, it's finesse, a new respect ...*

- *I was just touching everything with such care. You felt that the experts understood, and we weren't experts. [...] And then I kind of realized, and I was like okay … I can do it. I can see through a microscope. Even though it's a very simple step, you don't want to mess up anything that could make this be an error or complicate things or give misinformation that would be counter productive.*

- *I, too, didn't expect to look under microscopes and more in depth and learning about DNA.*

Evidence of Learning: Strand 4

Perhaps because the first, longer workshop involved participants who were already familiar with scientists and scientific work, we did not see much evidence in that workshop on involvement with the concerns of *LSIE* Strand 4:

- **Strand 4:** Reflect on science as a way of knowing; on processes, concepts, and institutions of science; and on their own process of learning about phenomena.

But in this second workshop, which attracted participants who were not familiar with science, there was strong evidence that some participants left the workshop thinking in new ways about science and scientists. Again, the evidence is in the participants' own words.

- *I don't have a scientific background. And so, somebody like me being invited to come here … I thought it was going to be well … okay … I'm going to be sent out to do some collecting and bring it back, and it was going to be pretty simple. I was a little surprised that the information that we got was in depth as it was. I'm glad for that because it just seemed more respectful of the people who were doing this. And, it made it very interesting. [This quotation is from one of the participants who attended out of an interest in doing some kayaking.]*

- *I mean … this has been very eye opening and just very informative about what you guys do here and what it takes to put a project like this together. We get to see you interacting with one another ... and the scientific work ... and trusting us ... and being very open and frank in front of us.*

- *Some of the DNA stuff was interesting. I didn't know that something as simple as something on ... like ... a blade of eelgrass takes that much ... that complexity, and that shear manpower to sequence and upload and compare. It seems like daunting task if we are to sequence everything that exists out there among us. So, that was kind of interesting to look at just the complexity of the process.*

- *I have a little bit of sensitivity to the suffering of other beings even though they're snails. And that I was very worried about. Even when I was collecting them, I was about to not grab a crab, because like I didn't know they were going to be killed. So, that part I didn't expect to think or feel that. But, even when I was handling my … the ones that were in the tupperware, I asked should I take them back to the sea, or will you do it, or what will happen to them?*

- *[I got] an exposure to scientists in a way that … when I heard them joke, or ... it was different. It was such a comfortable environment to learn. Even though I don't understand this completely and the names of the different organisms I was looking at … was something that I wasn't familiar with, it was like okay that's fine. We're all here together. They're trying to help us. That was very new to me.*

- *It's not like a science TV show where they're very specific about the terms or their normative way of doing things. It was so relaxed. It was like 'oh, look at this cool thing'. Or, it was … they say, 'Cool, it's moving and it's eating.' And it was so, so . . .*

- *I'm impressed by how much they seem to enjoy their work. Actually, it did rub off. And before I knew, I was looking around, but it wasn't until I saw, was it [scientist's name], said ... turning over rocks and saying oh look. And I thought, oh yeah why don't I turn over rocks too. Maybe I'll find something. And all of a sudden, it became more of a game, and it was really fun. Most of the afternoon was like that too.*

Participation in the workshop also stimulated some broader thinking about living systems and how people interact with them, as evidenced by these comments by two of the participants.

- ***First participant**: You know, you go in there, you disrupt an area, you know? And then you see what the fisherman ... you know ... you read in papers of fishermen who might try to, you know … and all that. And then here we are ... are we doing any ... you know what I mean? Tramping around in there picking stuff up. I mean ... I know we're not doing nearly the damage the local fishermen are. ... But, they got eat too, so I understand their plight. And, so ...*

- ***Second participant**: After seeing them under the microscope, it just seems more real to me.*

- ***First participant:** Yeah.*

- ***Second participant:** It seems 'real' to me for some reason. I don't think about these things when I shop ... or going on about my life. ... But, seeing it under a microscope, it was like ... you know ... I just suddenly felt connected to it and, you know, I felt ... uh ... kind of ... I don't know … small I guess.*

Evidence of Engagement: Strand 1 and Observation Protocol Development

I used this second workshop as an opportunity to begin testing and refining a protocol for collecting observation data about participant engagement. The paper included in Appendix C provides a detailed description of this observation protocol, along with the data collected during specimen sorting during this workshop.

Briefly summarized, the observation protocol seeks to divide observed behaviors into three categories:

- "Initiation" behaviors that are associated learning about the task to be completed and how to use instruments, process and preserve samples, and so on.
- "Doing" behaviors associated with the intended task.
- "Breakthrough" behaviors that are taken as evidence that a participant Is particularly interested in something that he or she has seen, is making a connection to past experience, or is in some other way moving beyond "doing" to expressing some personal connection to the activity.

As described in Appendix C, the observations during this workshop suggested that participants' experience in the workshop moved from initiation activities to doing, as one would expect. Once "doing" was underway, individual participants would periodically do something that we classified as a "breakthrough" behavior. What emerged as potentially interesting and useful was the aggregate pattern of behaviors over time, as illustrated in Figure 1.

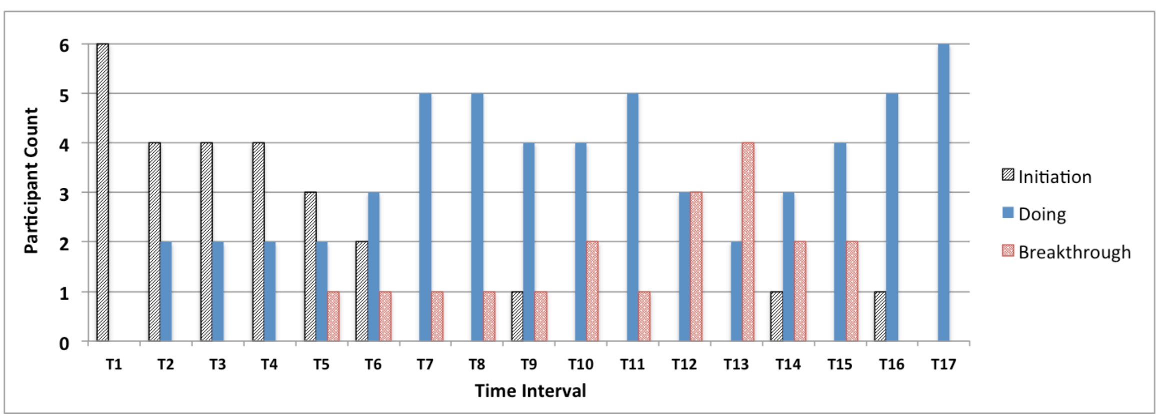

Figure 1. Time series of counts of different categories of behaviors during specimen sorting in the second eelgrass workshop. Time intervals are five minutes.

As the figure illustrates, breakthrough behaviors picked up about an hour into the specimen sorting work, then tapered off to be replaced by a complete focus on doing toward the end. Numerous explanations are possible. For example, perhaps the strong focus on doing at the end was related to the participants' awareness that there was only so much time for specimen sorting coupled with a desire to get as much work done as possible. At this early stage of work, my focus was not on trying to understand just this particular event, but was instead directed at using observation as way to characterize participant activities so that, over many events, patterns might emerge that could be useful in deepening understanding.

The observations collected in this event and subsequent events, as described in Appendix C, suggest that, with refinement, collecting such observations and undertaking such analysis might be fruitful.

Reflections on the Shorter Field Experience

This second eelgrass workshop suggests a number of conjectures that might be investigated in subsequent events.

- The shorter duration, coupled with a no-cost, high value activity like kayaking, did succeed in drawing participation from people with a pre-existing, deep interest in science.

- Once involved, these same participants appear to have come away with learning and experience that is related to science.

- The interaction with working scientists stimulated new thinking about what scientists do and about the kind of people that they are.

- As in the first workshop, once participants were present and engaged, the desire to contribute usefully to the scientists' work was "front-of mind" for the participants. The implication is that the connection between the participant's activities and the contribution that those activities make should be clear to participants.

In addition to the value that a kayak expedition adds as a way to attract participants, it also adds complexity. One of the participants capsized early in the trip. Getting the participant back into his boat and keeping him warm added to the demands on project leaders. It also illustrates the need to add capacity and capability to ensure participants' safety when kayaking of other more adventurous activities are part of the program.

it is important to note that reducing the time commitment and adding kayaking did not, in themselves, address the participant recruitment issue. The number of participants was still small. However, there are many businesses in the area that organize one and two-day expeditions for visitors to Acadia. These businesses succeed in fully booking their events—events that visitors pay to join. We conjecture that the issue in the case of the Biotrails Pathways program is that it did not have an established infrastructure for promoting and organizing such events. Given such an infrastructure, recruiting participants to short duration events seems possible.

Foregrounding the DNA Analysis

As the Pathway to Biotrails project progressed and as the project's efforts to establish a DNA barcode library for local species had continued success, we were increasingly sure that DNA barcoding could be a useful supporting technology for citizen science programs that depend on species identification. However, the Biotrails program also sought to identify ways in which DNA barcoding might be in the foreground of the learning, rather than just in the background as a key enabling technology.

The initial, two-weekend workshop foregrounded DNA barcoding in a number of ways. It involved participants directly in DNA extraction and creating gels. It also involved them in trimming and cleaning up the sequences for the DNA they had extracted and in searching for the sequences in the Barcode of Life Database (BOLD).

As described above, we determined that involving participants directly in DNA extraction was unproductive. However, we also spent some time in both of the eelgrass workshops showing participants about BOLD and how one could search for a match on a particular fragment of a DNA sequence. Participants found that interesting. So, we decided to build a one-hour workshop just around that part of the process. We offered the workshop as part of the 2014 Acadia National Park bioblitz.

This was yet another exploration of potential models for participant engagement. We wanted to see how hard or easy it was for volunteers to identify a useful segment of a digital representation of a DNA sequence and then use that segment to query BOLD. If it was easy enough, it could be possible to have volunteers do this strictly as an Internet-based citizen science activity, without even needing to be physically present at Acadia.

The event also provided another opportunity to test and refine the participant engagement protocol.

The DNA Barcode Lookup Workshop

Twenty-four volunteers participated in the activity. This group included approximately a half dozen students who were on the Schoodic Institute campus as part of an Earthwatch program along with their program leader. These students were already familiar with what a DNA sequence is and with the lookup process. Over the course of the one-hour workshop these Earthwatch volunteers assisted other participants.

Figure 2 shows the distribution of the different categories of participant behavior that emerged from participant observation. These observation data suggest that, unlike the participants in the specimen sorting whose activities are portrayed in Figure 1, these participants never moved completely beyond initiation behaviors. Further, breakthrough behaviors were less prominent as a proportion of overall activity than in the earlier workshop. Notably, initiation behaviors generally continued to be more numerous than breakthrough behaviors over the entire workshop.

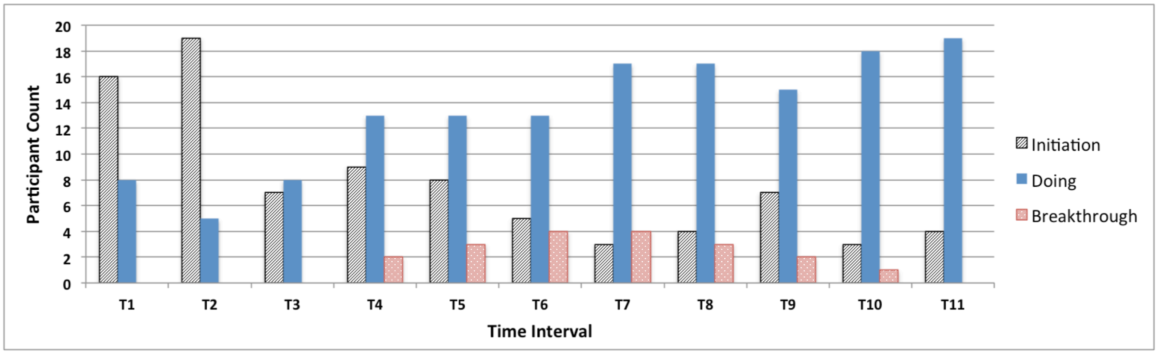

Figure 2. Time series of counts of different categories of behaviors during the DNA barcode cleanup and lookup workshop. Time intervals are five minutes.

These observation data are consistent with my informal impressions as I watched the workshop. The activity was challenging for a good number of participants. If the Earthwatch volunteers had not been present, scattered around the group and providing assistance and explanations, it would probably have been an even more challenging activity for the other volunteers.

Use of the Observation Protocol

As the comparison between Figure 1 and Figure 2 suggests, we found that the observation protocol was capable of portraying patterns of behavior for two different events that appeared to be potentially useful reflections of the events. We also found that even though the two events involved participants in substantially different kinds of activities, the protocol attends to behaviors that are general enough to be observable across these different events. The protocol needs additional development and refinement, as described in Appendix C and later in this report. The point here is simply that our experience with the observation protocol in the barcode lookup workshop was encouraging.

Bioblitz Observation

Since 2003, Acadia National Park has conducted an annual “bioblitz” to characterize biodiversity for a particular group of invertebrates. The bioblitzes attract experts from around the region, including professional entomologists and serious amateurs and hobbyists. These volunteers spread across the park and, over a period of 24 hours, intensively collect specimens in the target group. The collected specimens are used to document species occurrence and are useful in developing estimates of species richness.

These bioblitzes serve as an example of a kind of citizen science activity that might be enhanced through application of DNA barcoding to species identification. Consequently, they provide yet another opportunity to explore models for citizen scientist learning.

In 2014 the bioblitz focused on beetles. I observed a group of participants immediately after supper as they settled back into work on specimen identification and sorting. My goal, once again, was to test and refine the observation protocol. I was also interested in getting a quick, general sense of the kinds of interactions taking place between participants and of the learning opportunities that might be present.

Use of the Observation Protocol

Sorting took place in a science classroom at the Schoodic Education and Research Center. Participants were spread out at lab tables with microscopes and dishes containing specimens. They classified and in some cases killed, pinned, and mounted the specimens for display. As Figure 3 illustrates, initiation activity diminished and then disappeared over the first 35 minutes. This was not just because people became more expert. A majority of the participants had substantial expertise. But there were also a number of people present at the beginning—primarily younger people—who were not professional entomologists or expert amateurs. Many of these non-experts just watched what others were doing for a while and then left. The number of participants decreased over time from 17 to 9.

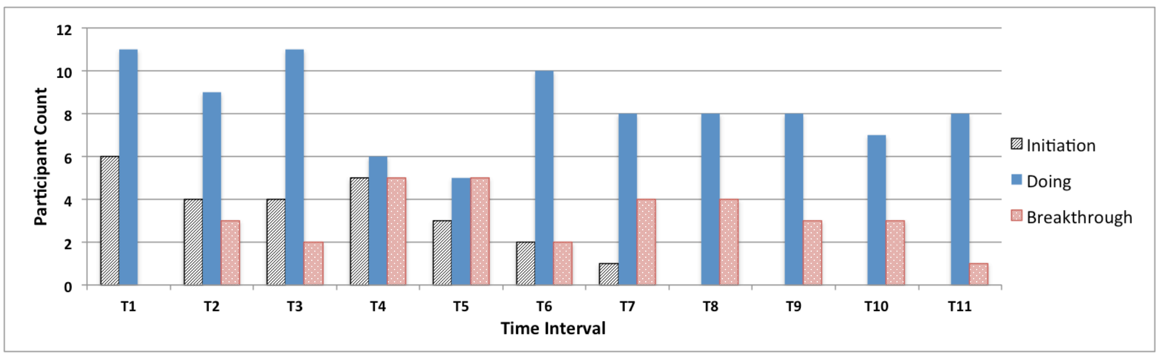

Figure 3. Time series of counts of different categories of behaviors during biotblitz specimen sorting. Time intervals are five minutes.

The coming and going of participants made recording and representing observations difficult. Additional work on the recording protocol would be necessary to adapt it for use with activities where participants arrive at different times and stay for different durations instead of participating as a group over a fixed period of time.

As noted in the paper in Appendix C, observing these participants also suggested a need to distinguish between different kinds of conversations among participants. An interaction where a participant finds something exciting and wants to share it with others is different in important ways from one where a participant asks another for help.

Other Observations About Participant Interactions

A bioblitz is a complicated event that includes many different kinds of activities. Consequently, my observation of classification and sorting activities should not be understood as reflection on the bioblitz as a whole.

What I saw was a group of professional and amateur experts who were primarily involved in work that was collaborative but also solitary. Many of the participants appeared to know each other. They worked quietly and steadily, periodically calling attention to an interesting beetle that others might want to look at.

They paid very little attention to newcomers who arrived to sit and watch. There was no obvious training, welcoming, or other induction process in place that helped newcomers move from watchers to doers. If a newcomer was competent enough to begin sorting and had a question, others would answer it. But the climb up from true "initiation" status to "doing" was sufficiently steep that a number of newcomers apparently decided not to attempt it.

This focus on getting a job done—sorting through literally thousands of specimens—seems well-suited to the goals of the bioblitz. But it is not clear to me that what I observed provided a learning opportunity that newcomers could negotiate. Again … what I saw might not be representative of other parts of the event.

Specimen Collection by Park Visitors

Several years ago, a biogeochemist working in association with the Schoodic Institute developed methods to use dragonfly larvae as bio-sentinels in studies of mercury pollution in watersheds (Nelson, et al., 2014). Subsequently, the National Park Service began using this procedure to monitor mercury in ponds, streams, and lakes across approximately 50 national parks. Some parks, including Acadia, use volunteers to collect dragonfly larvae as part of this monitoring program. We recognized this as yet another model of citizen science engagement that we could explore.

The Field Trip

Acadia posts notices of upcoming activities that visitors might join on a board at the park's primary visitor center in Hull's Cove. Dragonfly larvae collection field trips were listed along with other activities. I joined a park ranger on one of these trips. Two families, one with a ten-year old child and the other with two children, aged 5 and 11, joined us for the collecting work.

The ranger began with s brief talk about why the park was collecting dragonfly larvae and about mercury, where it came from, how it entered the food chain, and its effects on fish, birds, and humans. He showed the volunteers how to use a net to capture larvae and how to identify the larvae. He also demonstrated the collection protocol, emphasizing that it was important not to touch the dragonfly larvae because touching them might accidentally transfer mercury to the dragonfly and affect the analysis. He showed how to handle the dragonfly larvae with nets and plastic spoons. He gave a talk about safety, emphasizing that the mud at the edge of the pond was slippery. He helped volunteers put on boots and waders. The older child in the family with two children decided that she did not want to participate; everyone else started to work hunting for dragonfly larvae.

As the collecting proceeded, the ranger answered questions, periodically checked in on people, suggested that they try other locations, and helped with identification. But he was also responsible for ensuring that enough large larvae were collected, and so spent a good bit of time collecting on his own. The families worked as family groups; there was not much interaction between the families except as they brought specimens back to a common plastic collecting tub where they looked at what everyone had collected. Someone caught an eel; none of the participants had seen an eel. They spent a while watching it dart around in the tub. The five year old was completely absorbed in what he was finding in his net. Everyone seemed to have a good time; perhaps even a memorable time. The field trip collected the hoped-for number of dragonfly larvae, though most of them were smaller than the ideal specimen size. (Dragonfly larvae live for a number of years and collect more mercury as they get older and larger.)

Thoughts About Studying Learning in Such Events

Unlike the bioblitz, but like the other events that I observed, this field collection event had a clear beginning and end and involved the same group of people for the entire time. So, it fit with the observation protocol's recording model.

However, in this event it seemed to me that the really interesting learning opportunities emerged in the interaction between the parents and their children. This is consistent several decades of research focused on learning in museums (see, for example, (Dierking & Falk, 1994; Palmquist & Crowley, 2007a; 2007b). So, although the observation protocol "worked" in the sense that I was successful in recording a pattern of behaviors over the course of an hour, I felt that meaningful insight into the learning that was going required access richer data. Palmquist and Crowley (2007b) provide good examples of the kinds of insights that such data can provide, as well as a description of the difficulties inherent in collecting such data (2007a).

It is also important to note that collecting observation data in a field setting is relatively expensive. Unlike a museum setting, where the researchers can sit in one place while families come to them, observing fieldwork requires the observer to travel out to a remote site in order to watch a relatively small number of volunteers. This might be worthwhile if we were collecting rich data, such as audio recordings of conversations among family members, or if we were collecting observation data that could be considered together with other data collected over the course of a longer event. But when we are just collecting an hour or so of observations, it is possible that the expense involved in such data collection is greater than the data's research value.

I thought this was an interesting, rich event where different kinds of learning were going on. It seemed to me that variations in the nature of the activity, the way it was set up, and in how the ranger interacts with the participants would be likely to produce differences in the kind of learning that happens. As the museum research cited above indicates, the interactions within the family group are very important to learning. The family interactions can be responsive to the context created by the people leading the event. But research into such design and learning interactions would require collection of richer data than one can collect just through observation alone.

Reflection and Implications

The Pathway to Biotrails project began with a relatively rigid design for research on learning. It focused on questions about learning that related to one specific model for interacting with participants. We soon realized that this approach to learning about learning was inconsistent with the exploratory goals of the project. The overall goal of the Biotrails project was to:

> *Explore whether and how DNA-based identification of organisms, 'DNA barcoding', can extend and enhance informal science learning in citizen science projects involving adult volunteers. (from the first paragraph of the project proposal)*

The preceding pages document the exploration process that we followed in pursuit of this goal. I believe the evidence collected over the past two years suggests that the answer to the question of "whether" is "Yes" and that the answer to the question of "how" is much more complicated. In summarizing and reflecting on what I have reported here, I hope to unravel some of that complexity, ending finally with recommendations that will be useful in structuring future study of the learning that is enabled by DNA barcoding.

DNA Barcoding—Foreground or Background for Learning?

DNA barcoding makes it possible to identify organisms to the species level without access to taxonomic specialists. This project built a library of barcodes for local species and has demonstrated that the library is useful in biodiversity research and monitoring in the Acadia region. In other words, the Biotrails project has enhanced the potential for DNA Barcoding to serve as a key enabling technology that can support citizen science research.

It is in that sense that DNA barcoding provides "background" support for learning: it will enable more citizen science research that is dependent on species identification than would happen without DNA barcoding.

Over the past two years we explored the question of whether DNA barcoding can also usefully support learning in the "foreground," where the learners either engage directly in DNA barcoding or learn about it as a topic of interest.

In each of the eelgrass workshops and in the DNA barcode workshop we consistently collected evidence that volunteers are fascinated by DNA barcoding and enjoy learning more about it. But we also collected evidence that, without much more training and practice, volunteers are not able to make a useful contribution to the DNA barcoding process beyond initial species collection and sorting.

The evidence collected in this study suggests that DNA barcoding is more something that citizen scientists "learn about" rather than something they "learn to do."

Implications

The value of DNA barcoding to citizen science is in its ability to enable more citizen science research that depends on species identification. In the past, lack of access to taxonomic expertise has been a barrier to such work, particularly for local projects working on tight budgets.

Although there are still details to work out with regard to how local and lightly funded projects get access to assistance in extracting and sequencing DNA, it should be possible to create collaborative arrangements to provide such assistance. It should be possible to make access to DNA extraction and sequencing much more broadly available than is access to expert taxonomists.

One implication is that future projects might usefully focus on creation of and access to this kind of network that makes DNA barcoding capability available to citizen science projects.

With regard to learning, the implication is that the important questions regarding learning design and learning outcomes will be more related to the particulars of the citizen science work that DNA barcoding enables, rather than to DNA barcoding *per se*.

Reflections About Event Length

As we came to understand that DNA barcoding was more significant as an enabler of citizen science rather than as something that volunteers did themselves, we recognized that it would be possible to use DNA barcoding as the background support for a variety of ways to engage volunteers in citizen science, including events that were of short duration. Consequently, we included a number of short duration event structures in our exploration. We also developed an observation protocol that could be used in characterizing patterns of engagement in such events. I reflect on the observation protocol below.

Observing a variety of different kinds and lengths of events stimulated my thinking about the relationship between event length, learning, and research logistics. This thinking is not original; others have written about these matters. My purpose in sharing my thinking here is only to stimulate thinking in this report's readers, particularly in readers who are planning additional research.

From the standpoint of learning, deciding on the length of a citizen science event is a series of tradeoffs. The tradeoffs include matters such as the possibility of a deeper learning experience in a longer event that is counterbalanced by greater likelihood of attracting participants to shorter events.

From the standpoint of research on learning, I found that as events became very short and my interaction with participants decreased, the quality of the inferences about learning that I could make and support with evidence decreased.

The second eelgrass workshop, extending from late one afternoon through the next day, was long enough to provide good opportunities to collect evidence of participant's thinking and learning. The very short events were too short for more than formulation of general inferences from observation.

As noted earlier in this report, it is possible that collecting observation data from participants across a great many short events could result in useful patterns and variations across different event structures. But for some kinds of citizen science activities, such as field activities, the expense of undertaking such research could be prohibitive.

Implications

Researchers will be more likely to learn about the learning that is happening in citizen science events if the events last longer than just an hour or two, repeat over time with the same participants, or provide other opportunities for the researcher collect a variety of kinds of data about the participants' experiences. It is important to be able to talk with participants about what they are doing, in addition to observing them.

Reflections About Observation of Participant Engagement

The preceding sections of this report briefly describe an observation protocol that we developed to provide a consistent approach to collecting observations of participant behaviors. The report presents data collected through use of this protocol in its discussion of the different participant engagement models that we observed.

Appendix C provides a more detailed description of the observation protocol and argues that it would be useful to:

- Continue to refine the protocol, using larger numbers of observations of many more participants to support an empirical basis for grouping behaviors, as opposed to an *a priori* basis.
- Use the Rasch Model to transform the protocol's frequency counts to measurements that can be ordered on an interval scale
- Given such refinements, use the resulting instrument to explore design questions.

I will elaborate briefly on these ideas.

Continued Refinement

As the examples in this report indicate, the data from the observation protocol, when displayed as a time series of counts of different categories of behavior, appears to be usefully related to other observed characteristics of the event. Further, the protocol improved over time. Using it across events where participants were doing different things enabled me to expand and refine the list of behaviors included in the protocol. But it is likely that the current list of behaviors and grouping of behaviors need additional refinement:

- There are probably additional behaviors that should be included in the list, or general behaviors that are currently recorded (e.g., "sharing") that should be separated into a new set of distinct, differentiated behaviors.

- There may be behaviors that are not adding value to the observations but that are inflating counts of certain categories. It is important to identify such behaviors and to not record them.

- The grouping of behaviors currently just reflects my judgment about inferences that could be made from behaviors. This is the area where refinement is most important to establishing validity of inferences.

One way to address this last point would be to include a greater number of people in the evaluation and refinement of the protocol design. It might also be possible to cluster the behaviors empirically, using methods that I describe in the paper in Appendix C. Rather than using an *a priori* grouping of behaviors, one could use the Rasch model to generate estimates of "engagement effort" for each individual behavior. If engagement really does proceed in a step-wise fashion in which there are meaningful differences between high-engagement behaviors, routine "doing," and initiation behaviors, one would expect to see a kind of clustering of behaviors in terms of engagement effort. The clustering could provide an empirical basis for grouping behaviors.

Rasch Model

The paper included in Appendix C describes additional exploration of the observation protocol that I undertook during this project. Specifically, I explored the question of whether the counts of different behaviors could be transformed into "interval scale" measurements that can be added, divided, and used in statistical analysis in other ways. The paper provides evidence that the interval scale measurements resulting from applying a Rasch model to the data successfully capture the same characterization of the event that is portrayed by time-series frequency counts.

After presenting this paper at the Citizen Science Association conference earlier this year, a number of people asked for copies, which I provided with the understanding that they would send me some feedback. One thing that a number of readers asked as they corresponded with me was, essentially, "If the Rasch Model just provides the same picture as the frequency counts, what is the point?"

This was an example of my being so caught up in the analysis that I failed to communicate the bigger picture.

The point of converting the observations to interval scale data is not just to provide the same picture as the frequency count data (though that WAS important to demonstrating that the conversion did not fundamentally alter the data). The point is that, once we have interval scale data, we can compute means and, in other ways, aggregate the data.

Suppose, for example, that we wanted to investigate the question of whether participant engagement is different if leaders actively engage participants in constructing their own understanding rather than just telling them what is happening. If we had data from a number of events that were similar, but differed with regard to leaders' approach to supporting participant learning, the Rasch conversion to an interval scale would allow us to develop a quantitative investigation of the question.

Implications: Design Research

As noted above, DNA barcoding is not strongly associated just with one model of participant engagement and citizen science learning. It can support a wide variety of designs for learning. The work on an observation protocol begun here is still early-stage work. But, in demonstrating that such observations might, through Rasch modeling, ultimately result in measures that can added, divided, and averaged, we have taken a first step toward building a tool that could support comparisons between different designs.

Reflections About Interactions Between Leaders and Learners

Learning From Experts

Citizen science sometimes thinks in terms of program participants learning from experts. The volunteers do useful work, and as they do this work in association with scientists and other experts, they learn something about eelgrass, watersheds, astronomy, and so on.

My research apart from the Biotrails program focuses principally on studying teachers and learners in schools. Teaching is a conscious activity. As David Cohen (Cohen, 2011) puts it, there is nothing "natural" about teaching. It is an intentional act and something that people learn to do, and to do better.

In observing the different citizen science activities included in this report, it was only rarely that I saw evidence of conscious teaching. More often, scientists and expert participants were just doing what they do as scientists and experts and were not consciously thinking about the participants as learners.

This is not necessarily a bad thing. As some of the quotations from participants that I have included in this report indicate, some participants enjoyed and learned from the opportunity to see scientists being themselves.

But in other cases it seemed to me that not thinking about participants as learners was a missed opportunity.

One of the most striking differences between what I observed in this project and what I see in a good classroom has to do with the model of how people learn. In a good science classroom, learners are actively engaged in thinking things through, working with each other and sometimes disagreeing with and learning from each other as they encounter ideas, engage in scientific practices, and figure things out.

As described in the guidebook, *Ready, Set, Science!: Putting Research to Work in K-8 Science Classrooms* (Michaels, Shouse, & Schweingruber, 2008), the teacher in such classrooms works from a learning model that starts with the understandings that students bring to the class. The teacher seeks to build on, or in some cases, to change and improve these understandings. In order to do that, the teacher manages and guides the students' discourse—among themselves and with the teacher—keeping the focus on students' thinking rather than on what the teacher has to say. The goal is to help learners construct their own understanding, rather than on simply telling them something.

In their study of professional science interpreters in the informal science learning setting of a botanical garden, Zhai and Dillon (2014) found that most of the time these experts transmitted information authoritatively, providing little or no support for dialogue with the learners, much less among the learners.

I made similar observations across the events that I studied as part of this project. I encountered only one expert who responded to the participants' questions with other questions, encouraging participants to think about their questions on their own and providing guidance in the form of suggestions such as, "Well, you might look more closely at its head, and then see what you think." This expert had learned to move between her role as a scientist who possesses authoritative knowledge and her role as a teacher interested in helping learners see that they can figure something out on their own. In almost all other cases the experts just answered questions outright. In a few cases the scientist responded by saying that the question was complicated and that there was not time to answer it.

Mutual Learning

Citizen science should not *always* be about a transmission that runs from experts to learners. McCallie et al. (2009) argue strongly for thinking in terms of mutual learning through "public engagement with science." They write:

> *Public Engagement with Science literature and practice has a specific meaning that is characterized by mutual learning by publics and scientists—and, in some cases, policy makers. This orientation contrasts with a one-way transmission of knowledge from "experts" to publics. Specifically, PES experiences allow people with varied backgrounds and scientific expertise to articulate and contribute their perspectives, ideas, knowledge, and values in response to scientific questions or science-related controversies. PES thus is framed as a multi-directional dialogue among people that allows all the participants to learn. (p. 12)*

This more expansive conception of learning does not rule out instances of teaching, but it also clearly aims at something more democratic than what happens in classrooms. It is a model of learning that is directed at enabling learning among peers who have different kinds of knowledge.

The important point in terms of this exploratory project was that, just as with conscious teaching, I also did not find evidence of this kind of mutual learning in the programs that I observed. In each of the settings that I observed there was a clear distinction between experts, typically scientists but sometimes also serious amateurs, and other participants.

Implications

The future research projects that grow out of the Biotrails work should, first, be clear about the model of learning that the research is exploring. Is the learning model one that focuses primarily on how participants learn from experts? Or is it one that seeks to support mutual learning among peers, where everyone is a participant, a learner, and an expert?

The point here is that, in either model, research should attend to the nature of the interactions between experts and learners, even if those roles are fluid. In many studies this will mean paying particular attention to the forms and detailed structure of the discourse between participants, learners, and experts.

Reflections About Learning Research

Earlier, on page 26, I reflected on the connection between DNA barcoding and citizen science learning. I wrote that "the important questions regarding learning design and learning outcomes will be more related to the particulars of the citizen science work that DNA barcoding enables, rather than to DNA barcoding *per se*."

Put another way, there is no single, inevitable or necessary connection between DNA barcoding and learning. There is no 1-to-1 relationship between DNA barcoding and some particular approach to learning. Instead, DNA barcoding will have the effect of enabling many new and different citizen science investigations that focus on biodiversity. The quality and amount of learning associated with those investigations might be good or bad, large or small. The learning does not depend on the DNA barcoding. It depends on the learning design that is incorporated in each different project.

This is NOT the same as saying that the rapid maturation of DNA barcoding technologies has no impact on learning. Increased availability and decreased cost of DNA barcoding *will* impact learning because it will enable more biodiversity research. The important question, then, is how to support learning in biodiversity research. The DNA barcoding is in the background; the biodiversity research is in the foreground.

Implications

This is a big question. It will not be answered in a single study. But it should be possible to develop a body of research—multiple studies, not just one—that helps us understand the nature of effective learning in biodiversity research. This will be particularly true if the studies are set up to share and learn from each other's findings.

Keeping with the Pathway to Biotrails project's focus on exploratory work that will support more ambitious, focused research in the coming years, here are a few thoughts about what might make the studies in this corpus of future research most useful.

- **Goals for Biodiversity Learning:** The learning research will be productive only to the extent that there is clarity about what it is that is to be learned. As citizen projects, assisted by DNA barcoding, focus on biodiversity, what do we hope that participants will learn?

  There are, of course, different layers of answers to this question. At a high level, an answer might distinguish between learning from experts and mutual learning, as described above. At a lower level, the learning goals might be defined in terms of what the Next Generation Science Standards call "Disciplinary Core Ideas" that we would like people to understand. For example we might want to know whether some particular approach to organizing citizen science work makes it more likely that participants are able to think about the relationship between biodiversity and the resilience of a system when it is under stress.

  In short, the research should start with well-articulated conjectures about what will be learned and who will learn it.

- **A Focus on Clearly Stated Learning Goals:** This is really just a refinement of the previous point, but it is an important one. I have participated in a number of projects that frame learning in terms of constructs such as "attitude" or "motivation." These projects typically stumble because there is little agreement on what such terms mean. As Blumer (1955) argued way back in the 1950s, concepts such as "attitude" are empirically ambiguous and deficient as scientific concepts that are useful in research or for theory construction. Such terms are more indications of an area of interest rather than as something that is empirically useful.

  To the extent that research is focused on conjectures that connect citizen science biodiversity research to broad concerns suggested by terms such as attitude or motivation, the research design should be deeply rooted in the existing literature in order to be as precise as possible about just what it is about "attitude" or "motivation" that citizen science and biodiversity research is conjectured to affect.

- **A Theory of Action:** The research should begin with a clear "theory of action" comprised of conjectures about how participants' experiences and actions in the course of biodiversity research enable the desired learning. What structures, qualities, supports, and so on need to be present for the citizen science event to have the desired effect? How do these structures, qualities, and so on interact to create the effect? What will happen if some are missing?

  This kind of theory of action is not something that the research seeks to prove as true or false, but instead serves to guide the research. It is typically an interconnected set of conjectures. The research collects empirical (but not necessarily quantitative) data to explore the conjectures and how they are related. The theory of action is modified and improved in the process. At the end, if the expected learning does develop, the research will be able to offer empirical evidence supporting an overall conjecture about HOW it happened.

If the learning does not develop or does not develop expected, the empirical work on the theory of action might suggest what the problem is.

- **Getting Inside the Black Box:** This is related to the previous point, but is important enough to get a bullet of its own. Some research focuses on recording inputs to the process (here is what we did) and the outputs (here is what happened), but cannot get inside the black box that sits between the inputs and outputs to assert WHY the outputs happened and to help understand what was important to making them happen.

  My experience is that it is incredibly easy to end up doing this kind of research, where the question of WHY is not investigated, even when you don't intend to. It is particularly easy when a project is strongly focused on building something—on implementation. The project does some stuff. It uses a variety measures to characterize the outcomes. But no one can say for sure just how that particular implementation worked to produce those outcomes and not others. Everyone involved in such implementations always has a lot of conjectures about why the project was successful (or not so successful). But since the research was not designed to probe inside the many black boxes in a complex implementation, no can really say why things turned out as they did.

- **Commitment to Sharing Tools and Findings:** As noted above, no one study is going to provide a complete picture of how citizen science biodiversity research, enabled by DNA barcoding, will enable new learning. This picture will emerge only over the course of many studies. To the extent that those studies use shared tools such as those developed by the DEVISE project at the Cornell Lab of Ornithology and the early-stage tool described in this report, it will be easier to compare findings and to support conclusions that span projects.

References

Berkowitz, A. (1997). A simple framework for considering the benefits of SSPs. *Berkowitz, a.R. (1997). a Simple Framework for Considering the Benefits of SSPs. National Conference on Student and Scientist Partnerships: Conference Report. Cambridge, MA: TERC. Pp. 38–41.*

Blumer, H. (1955). Attitudes and the Social Act. *Social Problems*, *3*(2), 59–65.

Cohen, D. K. (2011). Teaching and its Predicaments. Cambridge, MA: Harvard University Press.

Dierking, L. D., & Falk, J. H. (1994). Family behavior and learning in informal science settings: A review of the research. *Science Education*, *78*(1), 57–72.

McCallie, E., Bell, L., Lohwater, T., Falk, J. H., Lehr, J., Lewenstein, B., et al. (2009). *Many Experts, Many Audiences: Public Engagement with Science and Informal Science Education*. Washington, DC: Center for Advancement of Informal Science Education.

Michaels, S., Shouse, A. W., & Schweingruber, H. A. (2008). Ready, Set, Science!: Putting Research to Work in K-8 Science Classrooms. Washington, D.C.: National Academies Press.

National Research Council. (2009). Learning Science in Informal Environments: People, Places, and Pursuits. (P. Bell, B. Lewenstein, A. W. Shouse, & M. A. Feder, Eds.). Washington, DC: National Academies Press.

Nelson, S. J., Chen, C., Kahl, J. S., & Krabbenhoft, D. P. (2014). *Validating Landscape Models for Mercury in Northeast Lakes using Dragonfly Nymphs as Mercury Bio-sentinels*. University of Maine Office of Research and Sponsored Programs: Grant Reports. Paper 52.

Palmquist, S. D., & Crowley, K. (2007a). Studying dinosaur learning on an island of expertise. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video Research in the Learning Sciences* (pp. 271–286). Mahway, NJ: Routledge.

Palmquist, S., & Crowley, K. (2007b). From teachers to testers: How parents talk to novice and expert children in a natural history museum. *Science Education*, *91*(5), 783–804. http://doi.org/10.1002/sce.20215

Penuel, W. R., Bienkowski, M., Gallagher, L., Korbak, C., Sussex, W., Yamaguchi, R., & Fishman, B. J. (2006). GLOBE Year 10 evaluation: Into the next generation. Menlo Park CA: SRI International.

Zhai, J., & Dillon, J. (2014). Communicating science to students: Investigating professional botanic garden educators' talk during guided school visits. *Journal of Research in Science Teaching*, *51*(4), 407–429. http://doi.org/10.1002/tea.21143

Zoellick, B., Nelson, S. J., & Schauffler, M. (2012). Participatory science and education; bringing both views into focus. *Frontiers in Ecology and the Environment*, *10*(6), 310–313.

Zoellick, B., Webber, H., James, K., Miller-Rushing, A., & Marion, M. (2015). Measuring Participant Engagement in Short-Duration Citizen Science Events. Presented at the Citizen Science Association Conference, San Jose, CA. Feb. 11-12, 2015.

Appendix A

To: Karen, Abe
From: Bill, Hannah
Re: Brief Reflections on September Eel Grass Workshops
Date: 23 September 2013

(I am writing this thinking that Hannah might agree, and I am writing it drawing heavily upon conversations that she and I have had and memos that we have shared … but, in fact, I am also writing it so late at night that she should be sleeping. It seems like a useful thing for us to kick around on the phone tomorrow (Tuesday) morning. Hannah is free to say that she disagrees with some or all of this … all errors and mistakes are mine. BZ.)

The September workshop was useful in shaping thinking about what we should focus on for the remainder of this pathways grant. We are still analyzing the data that we collected in that workshop through both observations and surveys and so do not have final results to share. However, given our meetings this week, it seems useful to share some generalizations emerging from the observations.

Context

For us the important context is shaped by (1) making use of the potential innovation afforded by DNA barcoding to advance informal STEM learning, and (2) our interest in using the pathways grant to gather evidence and experience in support of a more ambitious AISL proposal if and when there is another AISL funding cycle in the 2015 timeframe.

Participants

The participants in this first workshop were unusually engaged in science coming into our workshop. Some were professional scientists. Others were science teachers. Others had undergraduate training in science and a desire to reengage in scientific work and/or were active participants in other kinds citizen science activities. In short, this is not a sample from our target population of people with only casual interest in science whom we would like to reach in order to strengthen those interests. This is not to say that this was not a great (and engaged and forgiving) group for us to work with at the outset. It was indeed that. However, next year, we need to learn how to engage other kids of participants if we hope to make the case that we are advancing informal STEM learning.

Participant Motivation

The participants were interested first in doing something useful and secondarily in taking home some new learning or experience. Jane did a wonderful job of setting up a compelling problem and in making the case that she needed help. To the extent that participants expressed frustration, it was primarily because they were not sure that their work was adequately contributing to new understanding and a potential solution. (For example, a number of participants expressed concern about their lack of facility in the DNA work and the consequent low quality/quantity of information resulting from their work.)

The Groups

The innovation being tested and investigated in this study is the use of DNA barcoding to advance informal STEM learning. In this workshop we divided participants into two groups after an initial period of specimen collection and coarse sorting. One group worked toward more complete specification using a tabular key with the help of taxonomists. The other group got experience in the mechanics of extracting and amplifying DNA and in using the result of sequencing to match against the Barcode of Life database.

The initial design of our investigation built around the idea that using DNA barcoding would reduce the amount of time spend in identification work and thereby free up time and participant energy for other kinds of work and other kinds of learning. (Which is why we put questions about understanding of ecology and systems into our probes of participant learning.)

The design of the workshop turned out to be inconsistent with the idea of freeing up time for other things. The participants involved in the DNA procedures needed just as long – or longer – to complete their work as the folks working with microscopes, keys, and taxonomists.

In future reports, after telephone conversations with participants, we will share more about the experiences of these two groups. However, even at this point, we feel confident in making a few assertions:

- The participants in the DNA procedures group worked mostly individually, after receiving instruction; the work in the morphology group tended to be more social, with people working in pairs and with the whole group moving around to see each other's specimens.
- Neither group addressed anything beyond species identification; there was no focus on ecology or systems of relationships in the eelgrass community.
- Both groups saw their work as vaguely complementary, though the exact nature of the complementary relationship was not quite clear.
- Both groups were concerned about producing results that Jane Disney could use in her research.
- Both groups expressed satisfaction with their experiences.

<u>Building on The Innovation</u>

In the broadest sense, the idea behind this pathways project is that the use of DNA barcoding in citizen science will advance informal STEM learning. In this workshop we saw that having participants engage in DNA-related procedures provided them with a different kind of learning experience (following a DNA process recipe rather than using a key to identify species), but it did not result in a more productive experience from the standpoint of serving scientific objectives. (Many of the gels that participants prepared failed.) Since the participants in this study were more scientifically literate and well trained than the general public, it seems likely that the utility of participants' work in DNA processing would decrease in terms of scientific value as we engaged even less scientifically knowledgeable participants.

Our perception is that most important element in participants' sense of satisfaction was their belief that the were engaged in useful work aimed at helping Jane with her research problem. It follows that the participants' low productivity in the mechanics of DNA processing is a problem not only for the scientists who partner with the program, but also for the participants themselves.

Focusing on the service dimension of the project and on the quality of the participants' contributions suggests a different way of valuing the potential contribution from DNA barcoding: Rather than thinking in terms of time savings that enables other kinds of participant engagement, we might think of it in terms of increased quality of results. This is attractive for the scientist collaborator because he or she ends up with more useful data, and could be attractive for the participants in terms of their contribution. To achieve this improvement of quality we would need to move the DNA processing "offsite" – to a process undertaken by highly trained professionals.

This means engaging the participants in other kinds of activities. What activities? That will perhaps require us to broaden our notion of participation.

Advancing Informal STEM Learning

We are also concerned that we find a way to use the DNA barcoding to advance informal STEM learning (while also getting good, useful results for participating research scientists). One conjecture is that this might take the form of involving many more people in the project. Rather than thinking in terms of a few workshops involving a few dozen people, perhaps we should be acquiring evidence and experience related to a program that could involve thousands of people.

After observing this past workshop we conjecture that there might be many different kinds of engagement – related to the broader notion of participation mentioned above. (ALL of the different activities in this workshop were engaging for participants, given their sense of participation in useful, important work.) For some (probably local) people, participation could be year after year and could involve thinking about the design of the study, the logistics of implementation, and the analysis of results in collaboration with the lead research scientist. At the other extreme, we might seek to involve visitors to Acadia in useful work that might be accomplished over a period of a few hours. This work might be in the field (e.g., supervised sample collection) or in the lab (e.g., sorting specimens from debris). We would seek to use DNA barcoding as tool to allow us to do truly useful work in many locations at many times while engaging participants that range from fully collaborative work, on the one hand, to something more like a museum visit, on the other.

In order to undertake the learning and to collect evidence required to create a credible proposal for work of such scope and innovation, we would need to rethink our programs for the coming summer. We would need to think in terms of pilot studies that would help us learn how to engage participants at such a variety of levels. One important component would be establishing that we have ways to find and engage those different kinds of participants. Another would be, once we have engaged them, characterizing the quality and outcomes of those experiences.

In short, we think that the most important advances in informal STEM learning afforded by DNA barcoding might reach far beyond the more or less standard, long-form, deep engagement study that we outlined in our original proposal. We suggest that, instead, the project should think in terms of using the quality and quantity of results enabled by DNA barcoding (when implemented trained technicians) to open opportunities for useful, productive participation in scientific work for a much wider range of participants – varying in depth of knowledge and duration of engagement – than we have been able to reach to date. In the best case, we will find ways to ensure that engagement by such a diverse range of participants results in a more useful scientific outcome and a richer experience for participants at all levels of engagement.

Appendix B

To: Karen, Abe, Jane, Mark, Rick, Tina
From: Bill and Hannah
Re: Survey results from Eelgrass workshop
Date: 10 November 2013

The most important insights from our observation of the September workshop came from observations and interviews. An earlier memo that we wrote while our memories of the event were still fresh and which we sent to you some weeks ago summarized our analysis of those observations and interviews.

We also collected responses from two survey instruments at the outset of the workshop and again at the end. This memo is summarizes those participant responses. The actual survey instruments are attached to this memo.

Skills of Science

One of the instruments asks participants to assess their level of skill regarding a number of different kinds of scientific activity. The activities are grouped into four categories: inquiry skills, data-related skills, communications-related skills, and a last, mixed category of specialized skills. Figure 1 summarizes participant responses. This summary excludes responses from people who did not complete both the pre-event and post-event surveys.

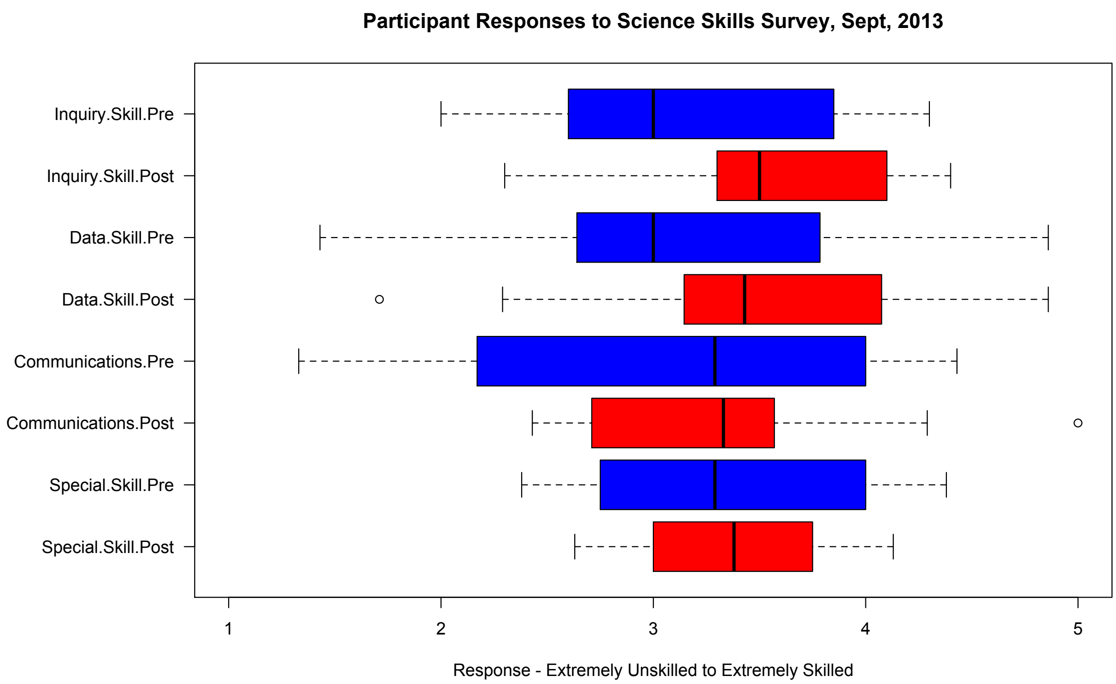

Figure 1. Self assessment of level of scientific skills

In this boxplot, the box shows the interquartile range for the middle quartiles, with a line indicating the position of the median. The whiskers show the range of responses outside of the interquartile range, up to a distance of 1.5 times the interquartile range. Any responses beyond that distance are represented as dots.

We observe that:

- Most of these participants feel that they are more skillful than not.
- It appears that, after the workshop, they felt better about their ability to engage in inquiry and to work with data. A paired sample, two-tail t-test shows that the change in perception of inquiry skills is significant ($p < 0.01$). (But, see the "Cautions" at the end of this report.)
- The variability in responses with regard to communication decreased.

Self-Efficacy for Science

The second survey asked participants to share their perceptions of their self-efficacy with regard to learning and understanding science topics and doing scientific activities. It also asks them how they feel about themselves as a scientist.

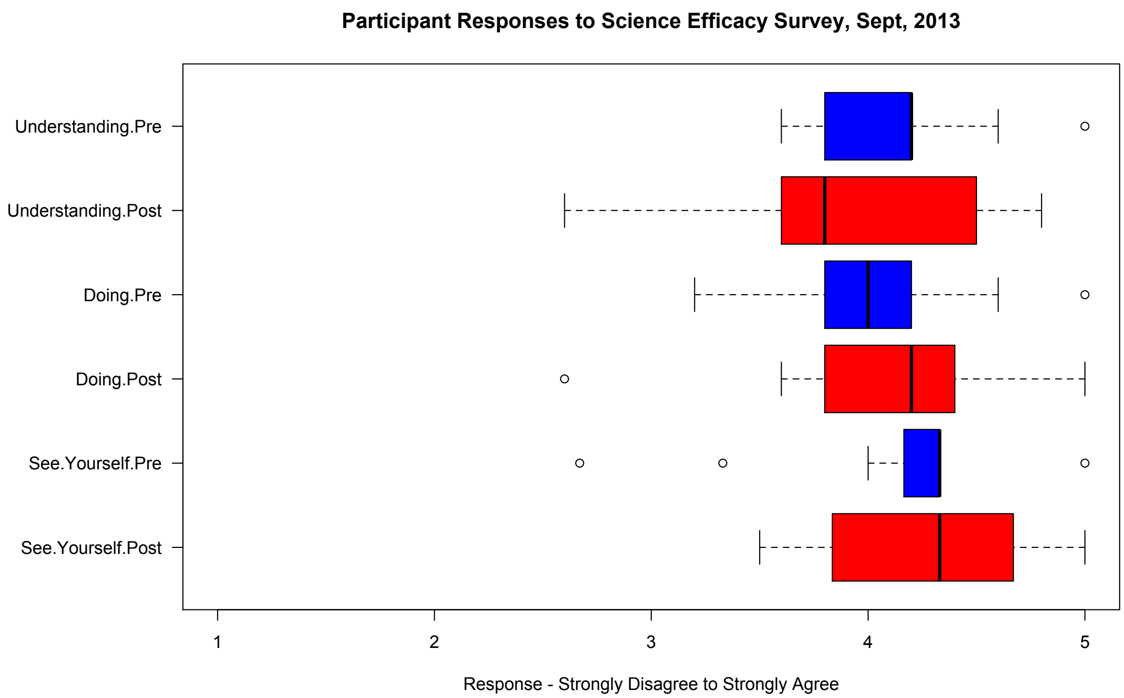

Figure 2. Perceived Self-Efficacy regarding science.

As in the responses regarding science-related skills, we note that this group is, in general, very comfortable with thinking about and doing science. None of the differences between pre-event and post-event results was statistically significant. However we note that the amount of variability in post-event responses is consistently higher than the variability in the pre-event responses. This suggests that the event may have made some of the respondents less confident of their science self-efficacy.

Summary

This group of participants came into the workshop with self-perceptions in the direction of relatively high science skill levels and relatively high comfort with and interest in science.

In general, differences between pre-event and post-event responses were modest, though there was a statistically significant difference in self-perception related to inquiry skills. As noted below under "Cautions," this is not a robust result.

Cautions

Apart from the finding that these participants tend to have a level of interest in science and comfort with science that is at the high end of the scale for these instruments, drawing other conclusions from these data should be done with caution. Since the responses in these surveys are recorded using a Likert scale, these measures are not on an interval scale. In other words, there is no reason to assume that the difference between a 3 response and a 4 response is the same as that between a 4 response and a 5 response. Adding up such values and then dividing to compute a mean will produce a result that includes error due to this scaling problem This means that a significance test between groups of Likert-scale scores is almost certain to include error that is not reflected in the t-test. Consequently, conclusions about "significant" differences are more suggestive than definitive.

Appendix C

Measuring Participant Engagement in Short-Duration Citizen Science Events

Bill Zoellick*[1], Hannah Webber[1], Karen James[2], Abraham Miller-Rushing[3], Michael Marion[2]

Collecting and analyzing observations of participants in citizen science programs is an important part of understanding and improving participant experiences and outcomes. This article describes an approach to designing and analyzing observation frameworks that has the potential to result in more useful, interval-scale estimates of the magnitude of theoretical constructs such as participant engagement, rather than simple counts of different kinds of behaviors. Specifically, the article explores use of a Rasch partial credit model in estimating participant engagement. Applying the model to two different sets of participant observations, it suggests that Rasch analysis has the potential to provide researchers and evaluators not only with statistically useful estimates, but also with insights into the performance of observation frameworks that can lead to framework improvement. The exploratory work described here suggests that larger scale testing and application of Rasch models to observations of citizen scientists is likely to be useful. It closes by outlining a program for the development and refinement of citizen science observation frameworks.

This paper explores a methodological problem. The problem emerged in preparing to investigate a conjecture about citizen science programs that are of relatively short duration and that take place in recreational settings.

The methodological problem arises as we seek to measure the degree to which participants are "engaged" in the citizen science work. As we explain later in more detail, collecting data about engagement in short duration events presents challenges that necessitate increased reliance on observations of participant behaviors, as opposed to collecting data through interviews and surveys.

Using observation data to construct inferences about the degree of participant engagement requires researchers to answer two measurement questions: (1) “Which behaviors should be recorded?” and (2) “What values should be assigned to these behaviors, corresponding to more or less engagement?” Taken together, the answers of these two questions comprise an "observation framework" that is used as a measurement tool in collecting data about engagement.

The methodological problem revolves around answering these two measurement questions in a way that considers them jointly and iteratively, using sets of recorded behaviors and initial assumptions about assigning values to behaviors to improve the measurement tool over time. A related problem is that of constructing the observation framework so that it provides interval scale measurements. Interval scale measurements are like those that we get from a thermometer, where the difference between 2 degrees and 4 degrees is the same amount of change as the difference between 4 degrees and 6 degrees. It would be very useful to have a way to measure engagement that produced interval scale values so that they can be added, averaged, and used in statistical tests to compare different program designs.

Describing the exploration of a methodological problem is not as straightforward as presenting the results of research using established methods. Rather than proceeding from research question to design to results and conclusions, this kind of paper is more tentative and, well … exploratory. Consequently, a brief overview of the structure of what follows will be useful for many readers.

We begin by describing why addressing this methodological question is potentially useful to any citizen science research that uses participant observation as an important source of data. We also provide a brief description of the research context in which this question emerged.

[1] Schoodic Institute at Acadia National Park; [2] MDI Biological Laboratory; [3] National Park Service

* Contact for corresponding author: bzoellick@schoodicinstitute.org

Zoellick, B., H. Webber, K. James, A. Miller-Rushing, M. Marion. (2015). *Measuring Participant Engagement in Short-Duration Citizen Science Events*. Paper presented at the Citizen Science Association Conference, San Jose, CA, Feb. 11-12, 2015.

Next, we look at an example of a simple observation framework that we adapted from research into the behaviors that people exhibit as they interact with exhibits in a science museum, which is another setting in which it is useful to understand participant engagement over relatively short time periods.

We then borrow an approach to measurement from psychometrics and education research known as the Rasch model in order to transform the results of the observations into interval scale data that we can graph and study. We describe the Rasch Model in enough detail to enable readers who are not familiar with it to follow as we move to the next step, in which we apply the Rasch model to two sets of observations of citizen science participants collected in different contexts.

We conclude with some thoughts about the potential value of what we learned through this exploration, along with the limitation of what we present here. Our goal is to describe what we learned with enough clarity that others might try the ideas presented here and either improve them or propose something better.

Motivation for the Work

The work that we describe here is part of research undertaken by MDI Biological Laboratory, Schoodic Institute, and the National Park Service that seeks better understanding of new citizen science opportunities that might be enabled by DNA barcoding, a technique that uses short DNA sequences to aid in species identification. The goal of this pilot-stage research is to identify, develop, and test techniques that might enable engagement in biodiversity-related citizen science by large numbers of park visitors and other people in similar settings. These pilot studies engaged participants in a variety of different activities, including sample collection, sorting of species in a lab, and work with computers to clean up and look up DNA sequences in online repositories such as the Barcode of Life Data System (BOLD).

The broader study explores several conjectures that reach beyond the work reported in this paper. One is that DNA barcoding can accurately identify invertebrate species of interest to biodiversity studies in the environs of Acadia National Park. A second conjecture is that having routine access to reliable species identification techniques will enable a broader range of citizen science activities, including activities that only require a few hours of a participant's time as they collect or sort samples. A third conjecture is that these shorter duration citizen science activities will enable engagement by participants who do not think of themselves as being deeply interested in science, but who are open to trying something new that provides a useful service.

It is this last conjecture that focused our attention on the problem of making sense of observations of participants in short-duration citizen science programs, which is the subject of this paper.

The Short Duration Problem

The approaches available to researchers describing participant engagement in citizen science depend on the nature of the citizen science activity. If a project engages participants over time periods spanning from days to weeks to months, interviews with participants are an excellent and often a preferred means to gain insight into participant experiences. If there are large numbers of participants engaged over such time periods, it often makes sense to use survey instruments to augment data collected through interviews.

However, not all citizen science work extends over periods of days, weeks, and months. Activities such as biodiversity inventories and other kinds of collection and sampling activities might involve participants for only a few hours. These kinds of sampling, data collection, and science support activities can be particularly attractive and useful in settings such as national parks, nature preserves, and other settings where people visit for short time periods and may be open to engagement in activities that they know are useful to the place and that offer opportunities for learning in addition to service.

Short duration events present a number of challenges to research or evaluation efforts. Since participants may only be involved for a few hours, asking them to complete a survey or interview can seem like a significant intrusion. Further, since participants in these free-choice activities are often on vacation or participating in recreational activities, being asked to complete surveys and interviews may be inconsistent with their motivation to participate in the activity.

The research and evaluation challenges that emerge in short-duration citizen science programs are not unique to citizen science. Researchers working in museum settings have confronted such problems for years. For example, Palmquist and Crowley (2007), in their article describing research on how parents and children interact with each other as they visit dinosaur exhibits, begin with a funny and sometimes painful account of the challenges encountered in presenting parents with an informed consent form to read and sign when they are holding the hand of a six-year old who is only 50 feet away from seeing the dinosaurs. In addressing such challenges, researchers and evaluators studying visitor engagement with museum exhibits often rely heavily on observing visitors, rather than interacting in more direct and intrusive ways.

Observation Frameworks

As Yogi Berra said, "You can observe a lot by watching." In the context of learning about participant engagement, observing a lot by watching implies having some way to organize what one sees as participants engage with a program or museum exhibit. Stated more formally, one needs an observation framework that directs attention to specific behaviors and that specifies how these behaviors are to be counted, aggregated or otherwise converted from simple counts into an assertion about what was happening over the course of the participation.

Observation frameworks can be loosely classified as either "high-inference" or "low-inference." High-inference frameworks provide guidance to researchers or evaluators with regard to evidence to be collected and considered, but analysis of the evidence and judgments about quality depend on the expertise of the observer. Horizon Research's *Inside the Classroom Observation and Analytic Protocol* (Horizon Research, 2000) is an example of a widely-used high-inference observation framework. It contains scales that the evaluator uses to rate the quality of lesson design, lesson implementation, lesson content, classroom culture, and a number of other characteristics. The final step in completing the protocol requires the evaluator to provide an overall "capsule rating" of the lesson, drawing on his or her experience in observing many lessons to render that judgment. High-inference frameworks can work well in a mature research area where there is substantial agreement about what constitutes quality.

Low-inference frameworks focus more on collecting data that are more directly connected to observation. They typically require the observer to record the presence or absence of certain behaviors repeatedly, at some fixed time interval. This kind of data collection still depends on inferences, but the inferences are implicit and embodied in the protocol designer's selection of behaviors that are assumed to be important and worth recording. The inference-making is done by the framework's designer, rather than by the observer who is on scene.

Because the low-inference observations collect data at a lower level of abstraction, recording behaviors rather than judgments about behaviors, they can be more useful and flexible in research that is trying to understand how things work, as opposed to making quality judgments about practices that are well-understood. This makes low-inference observation frameworks potentially attractive in relatively young research areas such as inquiry into the structure and effect of citizen science activities.

Adapting an Observation Framework for Use with Short-Duration Citizen Science

Chantal Barriault, working with David Pearson (Barriault & Pearson, 2010), developed an observation framework that serves as a good example of a low-inference framework as well as a starting point for the observation work described here. Working from many hours of observation as visitors engaged in activities in a science center, she theorized that visitors exhibit a sequence of three categories of behaviors as they interact with an exhibit. They begin with what she called Initiation activities consisting of behaviors such as looking at the exhibit, watching others interacting with the exhibit, and trying out the exhibit themselves. If the visitors are interested enough to continue interacting with the exhibit, they move to Transition behaviors that might include repeating the activity or expressing positive emotions in reaction to engagement. In some instances visitors move beyond the Transition behaviors to Breakthrough behaviors that might include referring to past experiences while they engage in the activity, seeking more information about the activity, sharing information with others, and engaging in inquisitive exploration about the activity through experimenting with it and trying out different kinds of actions. This pattern of engagement development is illustrated in Figure 1. Barriault and Pearson note that in their experience with using their framework in observing many visitors, not all visits fall into this pattern, but most do.

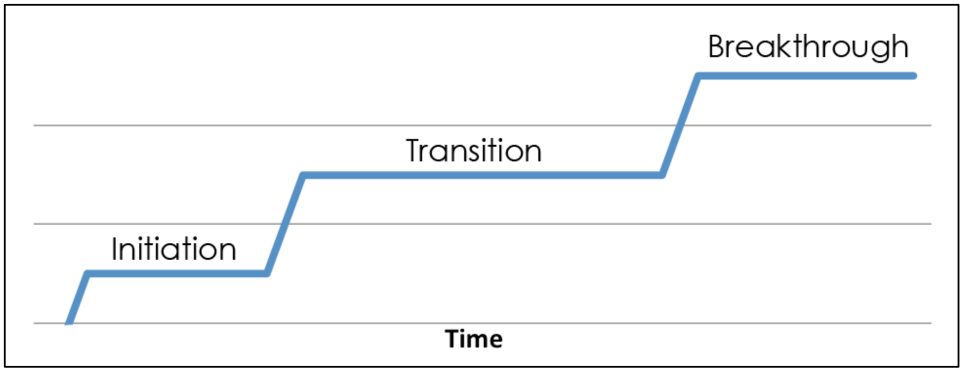

Figure 1. Expected pattern of behaviors in observation framework developed by Barriault and Pearson (2010).

Barriault and Pearson developed this framework in order to make decisions about exhibit design. Consequently, they were primarily interested in ensuring that an exhibit elicited a sufficient number of behaviors in the Transition or Breakthrough behavior categories. If observations revealed that visitors were only engaging in Initiation behaviors when interacting with an exhibit, that would suggest that the exhibit needed to be revised in some way.

We recognized that Barriault and Pearson's observation framework might be adapted to address the problem of making sense of observations of citizen science participants. It seemed reasonable to conjecture that participants would begin by focusing on mastery of the protocols for collecting, sorting, processing, or citizen science activities, and then, having mastered the fundamentals, would move on to higher levels of personal engagement with the activity.

Initial Shaping of An Observation Framework: What Behaviors Do We Record?

One of the key requirements for an observation framework is that it takes account of all the behaviors that seem meaningful and important. This is not just analytical work; it is also deeply empirical. If we are measuring engagement, one has to see what visitors or participants actually do, making sure that behaviors that appear to be important are recorded and included in the final assessment of engagement. Barriault (Barriault, 2014) provided a good example of this kind of work as she described the steps she took to adapt her initial observation framework, developed for use in science centers, for use in zoos and aquariums, where visitors were interacting with living animals rather than with more predictable exhibits that are designed by museum staff.

We used close observation of six citizen scientists who volunteered to assist in specimen sorting to begin the task of identifying the behaviors to include in an observation framework for the kinds of citizen science lab activities that might be part of biodiversity research programs. These six individuals provided informed consent to participate in human subject research, which meant that we could collect a variety of kinds of information, including interviews, that enabled us to connect the behaviors that we were observing with the participant's own perceptions of what they were doing. The laboratory tasks that the volunteers undertook included using dissecting microscopes to scan eelgrass blades to find and remove invertebrates living on the eelgrass. Then, working with the assistance of a taxonomist and from photographs of common invertebrates, the citizen scientists attempted to identify the animals that they found and placed them in containers for subsequent DNA analysis. We made observations at five-minute intervals over a period of 90 minutes.

One realization that emerged quickly was that rather than the simple progression from Initiation through Transition to Breakthrough illustrated in Figure 1, we were witnessing something more complicated. We did generally see evidence of Initiation, but the initiation was followed not by a Transition stage, but instead by a lot of "Doing" of the assigned task. Over the course of the Doing the participants periodically engaged in Breakthrough behaviors, but went back from Breakthrough to the ongoing "Doing" that was the primary focus of the lab work. Figure 2 illustrates this pattern.

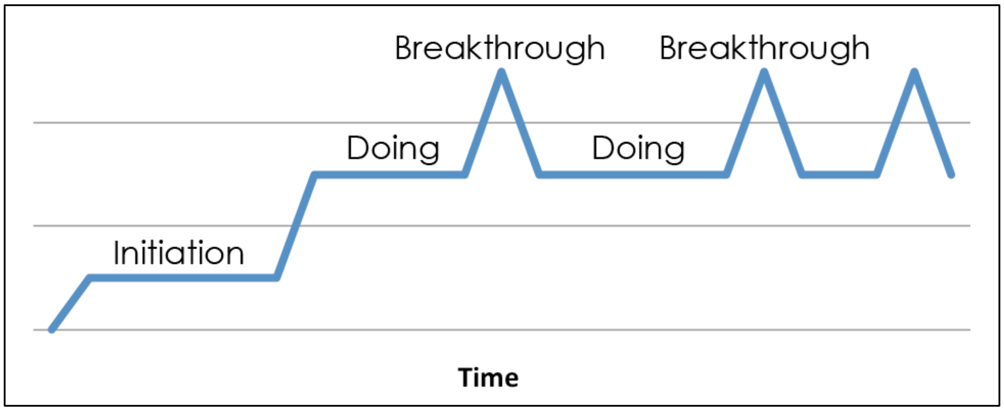

Figure 2. Pattern of behaviors observed in citizen science laboratory work.

Rather than a progression to successively higher levels of engagement in which the middle level is a Transition to something else, the middle level of engagement in this laboratory work appeared to be a kind of "base" state. Participant engagement might periodically rise above this base state, but generally returned to it within the five-minute observation interval. In retrospect, this finding makes intuitive sense: museum visits are primarily free choice activities in which visitors engage with an exhibit as long as it holds their attention, but then move on to something else. Citizen science activities, on the other hand, invite participant commitment to doing some work in support of research over a period of time that might extend for an hour, a day, or years. A participant's level of interest might wax or wane over this time period, but the work goes on.

Conversations with participants as they engaged in searching for invertebrates, sorting them, and preserving them contributed to a categorization of different behaviors as being either Initiation, Doing, or Breakthrough behaviors. For example, we were interested to note that participants took out smart phones to take pictures through their microscopes of what they were seeing. After talking with them about what they were doing and why, we decided to categorize this particular behavior as Breakthrough because it provided evidence of deeper participant engagement beyond the routine Doing of looking, sorting, and preserving. Table 1 provides a complete list of the behaviors that we recorded, together with the scheme for categorizing them as Initiation, Doing, or Breakthrough behaviors.

Table 1. Categorization of behaviors observed during citizen science laboratory activities

Initiation Behaviors	
Ask	Participant is asking questions about how to do things, about process, about protocols – seeking how to begin work
Try	Participant is interacting with equipment, tools, procedures in a tentative way to learn how things work
Watch	Participant is watching others use equipment, tools, procedures in a tentative way to learn how things work
Doing Behaviors	
Doing	Participant is engaged in intended task
Commenting	Participant is commenting on the process or task as he or she learns it. (e.g., "the focus knob is very delicate," "these things won't hold still")
Waiting	Participant is waiting for a resource, assistance, identification, etc.
Seeking Help	Participant is asking for assistance, clarification, additional information to improve at task
Breakthrough Behaviors	
Providing Help	Participant is providing assistance or information to another participant
Sharing	Participant is sharing information / insights / excitement with others ("Come look at this!") NOTE: "Sharing" is not the same as "chatter" that happens in the course of doing the activity. Sharing is related to something perceived, discovered, noticed, seen, and so on ... related to the work. Chatter and small talk in the course of doing the activity is just "Doing."
Pictures	Participant is taking pictures related to the task or subject matter (e.g., pictures of things collected, pictures of other participants, pictures of setting) NOTE: try to describe what is being photographed
Reference to past work	Participant is relating this experience to something he or she has done before. (This is a more specific kind of sharing … record as "R" rather than "S")
Inquiry	Participant is involved in manipulation of equipment, comparing of samples, and so on to achieve better understanding of the topic at hand.

Interpreting the Patterns

As Figure 1 and Figure 2 suggest, one approach to making sense of these observation data is to look at how the character of the behaviors changes over time. Figure 3 shows how the frequency of the different categories of behaviors changed over the hour and a half of work in the lab as participants inspected eelgrass for invertebrates. Initiation activities decreased over the course of the first half hour (up to interval T6). During this time the participants became familiar with using microscopes, sought more specific instruction, watched other participants, and waited for help. After interval T6, most of the participants were engaged in the intended task. After about an hour of work (at interval T12), one of the participants found a hydrozoan and another found a ribbon worm. This resulted in Breakthrough behaviors including sharing and taking pictures of the animals. These behaviors persisted for several more time intervals.

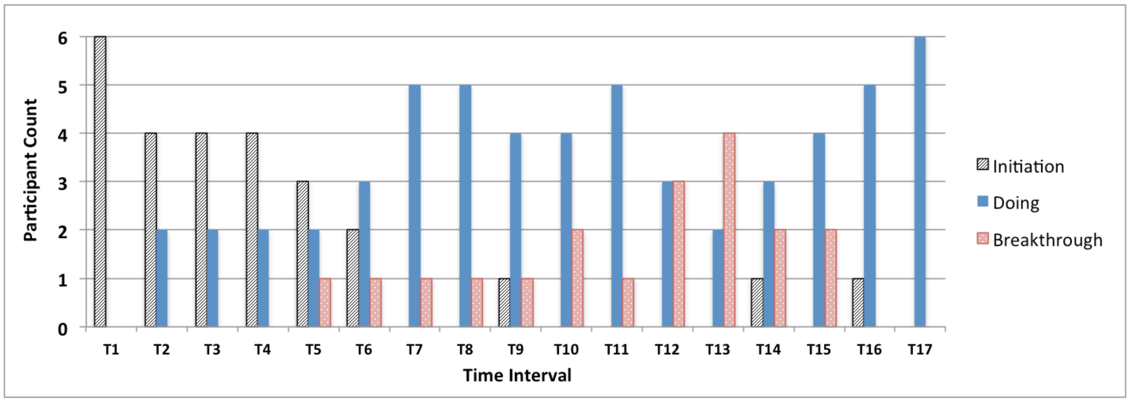

Figure 3. Frequency counts of different categories of behaviors during eelgrass lab activities.

The Rasch Model as a Way to Move Beyond Frequency Counts

As Figure 3 illustrates, looking at changes in frequency counts for different behavior categories over time helps make sense of observations. For studies where there are only a small number of participants, looking at patterns of frequency counts may be the most powerful analysis that the data can support. But studies that collect data about 50-100 participants, or more, have access to more powerful options.

The problem with frequency counts of different categories of behaviors is that they are counts of different kinds of things that are not related in any way that is obviously measurable. For example, in the simple observation framework presented above, we have defined Doing and Breakthrough behaviors as categorically different, but have no easy way to use the data we have collected to refine this definition. Perhaps some of the behaviors now labeled as Breakthrough should really be considered as Doing, or *vice versa*. It would be useful to have some way to put these different categories on a single scale that represented different degrees of engagement. Having such a scale would allow us to use statistical methods to explore such questions as, "Do the participants who are generally exhibiting higher levels of overall engagement consistently perform behaviors that we have placed within the Breakthrough category?" If the answer to such a question was "No," we might then consider revising our observation framework. Access to methods that would support iterative improvement of an observation framework as it was used with hundreds of participants across multiple studies would be beneficial to citizen science participant research as a whole, in addition to being useful within individual studies.

An additional benefit of being able to think of engagement as a scale of values, as opposed to counts of different categories of behaviors, is that it would support research questions that involved using inferential statistics to make comparisons between different sets of observations related to different program designs or different groups of participants.

The Rasch Model

Over the past four decades, psychometric research has made increasing use a model developed by Georg Rasch (Rasch, 1980) to address this kind of scaling problem. The Rasch Model treats the probability of responding correctly to a test item as a function of both the item difficulty and the ability of respondents who answer it correctly. Rather than assigning the same value to all items in computing a score (e.g.,, a value of 1 for a correct answer and 0 for a wrong answer), it computes a different difficulty level for each item. Computation of this difficulty metric reflects the overall scores of those who answer the item. So, a more difficult item is one that more capable respondents can usually answer and that is challenging for less capable respondents. Clearly, there is some circularity in this definition of difficulty, since we identify the more capable respondents on the basis of their scores, and those scores depend on the item difficulties, which, in turn, depend on knowing who is more capable. Consequently, producing a list of item difficulties is an optimization problem that is solved iteratively on a computer, over a series of successive approximations. But the modeling process converges for sufficiently large numbers of items and respondents. One of the important and useful features of the Rasch item difficulty levels is that they are interval scale measurements. In other words, the amount of ability, engagement, or whatever else we are measuring to move from a level of 1 to 2 is the same as the amount required to move from 2 to 3.

Rasch's original model worked only when answers were simply right or wrong, but it has now been extended to work with "partial credit models" in which the answer to an item can have a range of values that increase as answers demonstrate deeper understanding or, in our case, deeper engagement. Bond and Fox (2012) provide a good overview of Rasch models, including partial credit models, for readers seeking more detail.

The Rasch partial credit model is potentially applicable to the problem of assigning more meaningful and useful values to the behaviors observed during citizen science programs. We use the word "potentially" because the Rasch model requires adherence to some assumptions about the thing being measured. Specifically, it assumes that performance and ability are related along a single dimension. This assumption would be violated if we were measuring two different things at once, for instance speed and strength. Some highly capable respondents would do well on some difficult probes, but poorly on others. Rasch modeling would not fit that kind of measurement tool.

For the measurement problem explored here, meeting Rasch model requirements would mean that engagement, as defined by the behaviors in the observation framework, should emerge as a single construct. If we find that we can order participants in a consistent way from least engaged to most engaged and if we find that this ordering correlates with the degree to which engagement is more challenging at different times, we will have a measure for which Rasch modeling might be useful. On the other hand, if we uncover a more complicated relationship between participants and engagement, the Rasch model will not fit and—importantly—the Rasch analysis will tell us that. In short, trying to fit the Rasch model to our observation data can contribute to a better understanding of the construct that we are trying to measure.

Applying the Rasch Model to Samples of Citizen Science Observations

The observations of the six participants as they found and sorted invertebrates living on eelgrass gave us the first draft of an observation framework that, with refinement, might be useful in measuring engagement. We were now in a position to try the framework out on some larger groups of participants to see if we could use the Rasch model to begin a process of iteratively refining the framework. We collected two sets of observation data in two different settings.

Observing and Measuring Engagement During Computer Lab Work

One set of observation data was collected during a computer lab workshop in which participants worked to clean up digital DNA barcode records produced from local lab work. Once they had a clean record that could be used in a search against the BOLD barcode repository, they looked for a match on the record they had cleaned up. The workshop leader made it clear to participants that she was primarily interested in getting feedback from participants about how to organize this kind of activity so that it might be done remotely by other citizen science participants.

The workshop involved 24 participants for an hour. Behavior was observed every five minutes using the observation framework in Table 1. Figure 4 summarizes the frequency distribution of different classes of behavior over the course of the hour.

The Rasch model assumes that a measurement instrument consists of a set of items, or probes, that are each scored separately. In applying the model to this set of observations, we treated each time interval as a separate probe and looked at the distribution of behaviors for each interval. Rasch analysis software, by convention, uses a value of 0 for the lowest category of response. Consequently, we mapped the different categories of observed behaviors, Initiation, Doing, and Breakthrough, to category numbers of 0, 1, and 2.

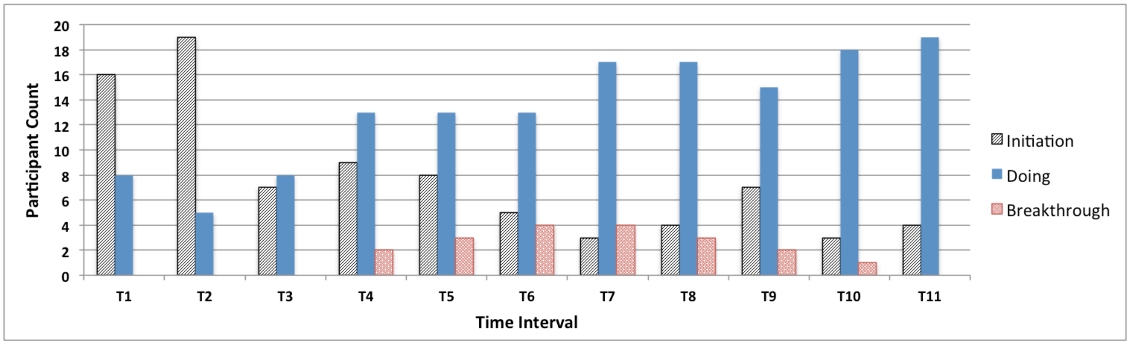

Figure 4. Frequency counts of different levels of behavior during DNA barcode clean up and look up.

The actual Rasch analysis of the observation data used the partial credit estimation in the 'eRm' package (Mair, et al., 2015), which is implemented in the R statistical programming language. Partial credit estimation works on the assumption that a transition from one category of response to another, say, from category 0 to category 1, is evidence that the respondent has moved beyond some threshold on the scale of values that is being constructed through the Rasch Model. In this case, that scale of values is an estimate of the amount of participant engagement. Figure 5 illustrates how this works. Working from the observation data, the analysis computes the probability of observing behaviors corresponding to the three categories at each time interval. In this figure we look just at intervals T4 and T6.

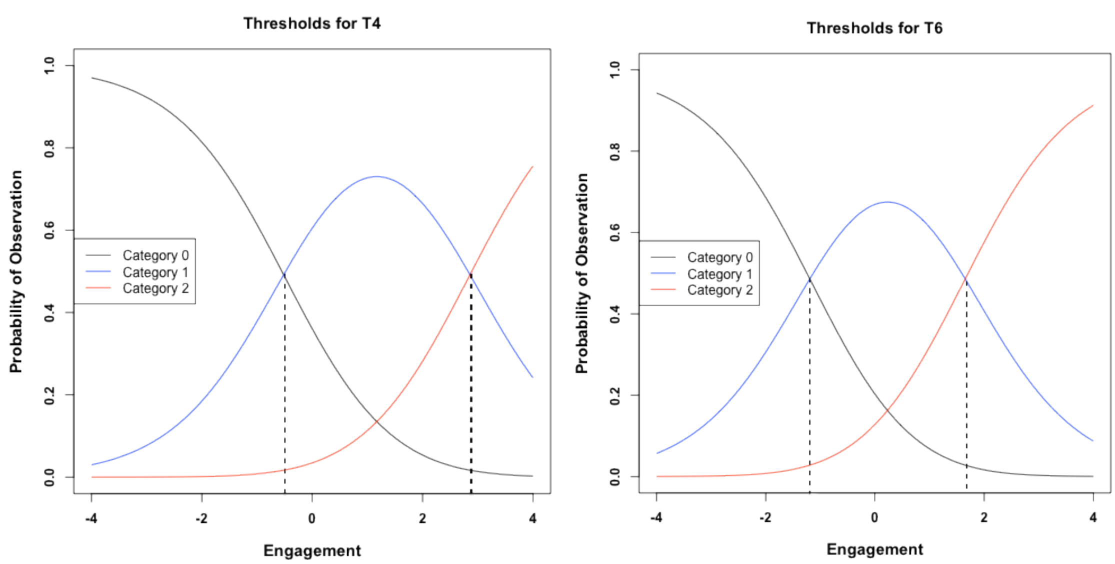

Figure 5. Converting categories of observations to levels of engagement.

The analysis computes the size of the intervals on the horizontal axis and the range of values that are displayed—in this case a scale that we are asserting to represent the degree of engagement—by considering the distribution of engagement across all the participants over the full duration of observations. The model is set up so that participants with an average level of engagement have a 0.5 probability of displaying an engagement level of zero. Note that this means that a negative value on this scale does not mean "nonengagement." The scale resulting from Rasch analysis is like the Fahrenheit or Celsius temperature scales, where the placement of zero is according to convention or design, rather than an indication of an absolute value.

The dotted lines that drop from the points of intersection between successive categories to the X-axis show how the model converts the probability of observing a particular category of behavior (a probability that is related to the frequency counts shown in Figure 4) into a measurement of engagement. This translation moves from something that can actually be observed (categories of behaviors) to a theoretical construct (in this case, engagement) that can only be inferred from the observations.

Comparison of the engagement values at the two time intervals shows that, according to this model, the level of engagement required to make the transition from Initiation (Category 0) to Doing (Category 1) at interval T4 was about -0.5. Ten minutes later, at interval T6, the dotted line drops down to the X-axis at a value of approximately -1.2. So, moving from Initiation to Doing required 0.7 less engagement at T6 than it did at T4. What is even more interesting is that the level of engagement required to transition from Doing (Category 1) to Breakthrough (Category 2) showed an even greater decrease over the same ten minutes. An engagement level of approximately 2.9 was required to move to Breakthrough at interval T4, but this decreased by 1.2 to a value of approximately 1.7 at interval T6. Because the Rasch model produces an interval scale of values, it is possible to compare these two amounts of change. We can say that it appears that the engagement required to move to Breakthrough decreased about twice as much as the level required to move from Initiation to Doing over this ten minute period. This is the kind of analysis that having an interval scale enables. Clearly, we would want to look at many more observations before making strong assertions. The point here is simply that such comparisons of values are not even possible when researchers can only look at frequency counts, rather than at an actual measurement scale.

Table 2. Engagement transition thresholds at different time intervals during DNA barcode work.

Interval	Threshold 0-1	Threshold 1-2	Interval	Threshold 0-1	Threshold 1-2
T1	0.62	not present	**T7**	-2.07	2.16
T2	1.43	not present	**T8**	-1.96	2.71
T3	0.16	not present	**T9**	-1.31	3.04
T4	-0.52	2.86	**T10**	-2.22	5.89
T5	-0.93	2.22	**T11**	-2.32	not present
T6	-1.19	1.65			

Table 2 summarizes the transition thresholds for all of the observation intervals during the DNA barcode clean up and lookup program. A scan of the table's columns reveals that the transition from Initiation to Doing became relatively less demanding over time. This is what we would expect as participants become more familiar with how to do the required work in a citizen science program. The pattern of thresholds for transitions from Doing to Breakthrough is more interesting. The estimated engagement effort required to transition to Breakthrough activities decreased for a few time periods, but then began to increase. Figure 6 shows this pattern graphically.

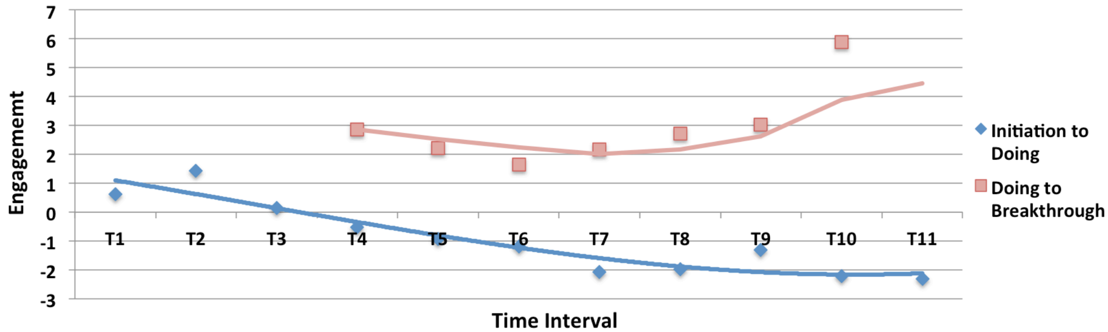

Figure 6. Amount of engagement required to transition between behaviors over time during DNA barcode work.

Applying the Rasch partial credit model to these observation data does not, of course, explain this initial decrease and subsequent increase in engagement associated with Breakthrough behaviors, it only assists in bringing the pattern into focus. We can usefully distinguish between two different kinds of conjectures about possible causes for the pattern. There are conjectures that assume that the pattern is correct, and look to program design, context, and participant motivation for explanations. An example of such a conjecture is that, perhaps, as the novelty of the activity wore off, participants were less engaged. Such conjectures are warranted when we have confidence in the measurement instrument. Making and testing such conjectures is the reason to have good instruments.

When we are using an observation framework that is still at such an early stage of development, there is good reason to question whether the measurement instrument is producing good estimates of engagement. This leads to a second kind of conjecture that considers the possibility that the observation framework needs refinement. For example, a review of the actual hand-coded observation record for the data presented above reveals that a substantial proportion of the behaviors that were classified as Breakthrough behaviors during this event were instances where a more knowledgeable participant was providing assistance to another participant. What difference would it make if we categorized these behaviors as instances of Doing rather than Breakthrough? Figure 7 shows the answer to this question.

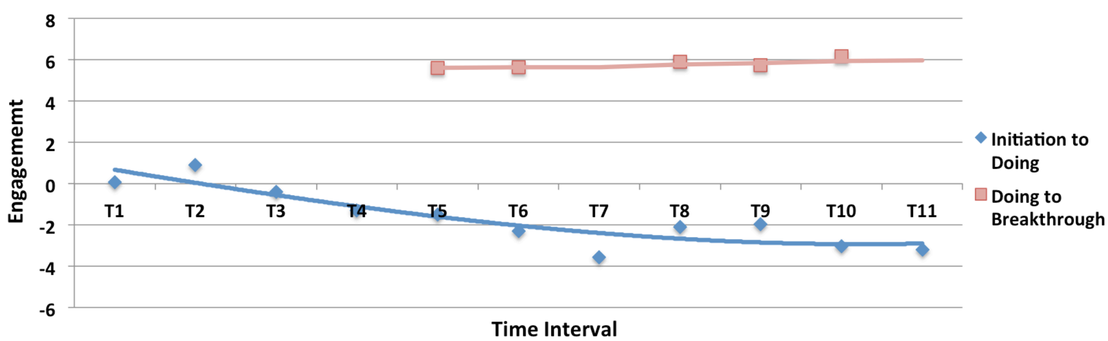

Figure 7. Change in engagement required for Breakthrough after adjusting observation framework: DNA barcode work.

The important point that this example illustrates is that by giving us a way to create a scale for measuring engagement, the Rasch analysis provides us with an easy way to see the effects of making a change in the observation framework. If we think of constructing an observation instrument for measuring engagement in the same way that one would approach constructing other psychometric measures, we would expect instrument construction to be iterative over many trials and to be concerned with validity: what is the chain of reasoning that connects evidence about the instrument's performance with the claim that it is measuring the theoretical construct of interest? Having the ability to convert counts of observations to an interval scale facilitates such inquiry. In this particular instance, the unexpected shape of Breakthrough engagement plot in Figure 6 raised the question of whether providing assistance to others is more properly just part of the job, instead of evidence of deeper engagement. This kind of question stimulates important thinking about just what we mean by "Breakthrough" and about the nature of engagement. Pursuing that line of thinking may require interviews with participants who were providing assistance to ask them questions about their experiences. The Rasch partial credit model provides quantitative input to a dialogue that is at once quantitative, qualitative, and epistemic and that is an essential part of instrument design and construction.

A Second Test Case

We used the same observation framework to collect data in a different setting with different kinds of participants. Since our focus is on addressing a measurement problem related to participant observation, there is value in applying the measurement tool and techniques in different observation contexts.

We collected observation data during a bioblitz event as participants sorted beetles that had been collected earlier that day. Observations took place in the evening immediately after participants had finished supper and had settled back to work. We observed 17 participants at five-minute intervals for an hour. The group included amateur and professional entomologists who were generally comfortable with identifying and preserving samples as well as people with less experience with insects. Figure 8 summarizes the distribution of observed behaviors.

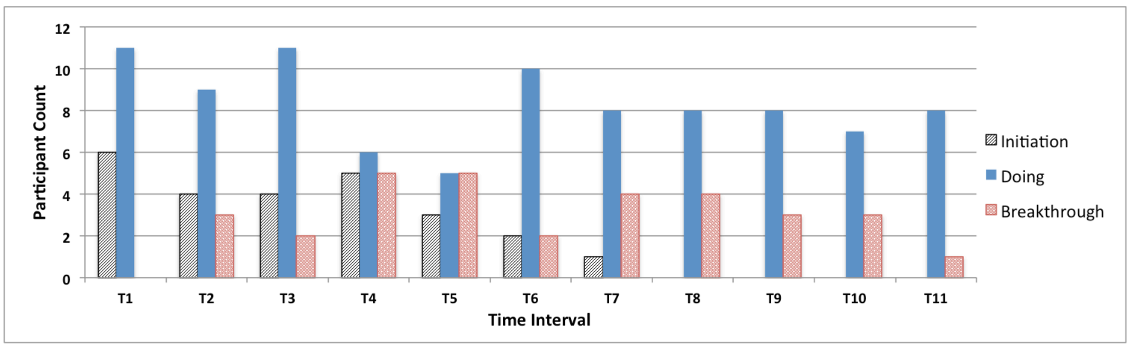

Figure 8. Frequency counts of different categories of behavior during biolblitz specimen sorting.

As one might expect, this group was substantially involved in Doing right from the outset. Participants who were not as familiar with the sorting and pinning process tended to begin by watching more experienced workers until they had a better idea of what was going on. In some cases, some "participants" never moved to the Doing stage, but just watched for a while and left.

Applying the Rasch partial credit model to these data, using the same tools and procedures as before, produced threshold data summarized in Table 3.

Table 3. Engagement transition thresholds at different time intervals during bioblitz sorting.

Interval	Threshold 0-1	Threshold 1-2	Interval	Threshold 0-1	Threshold 1-2
T1	-0.76	not present	T7	-3.54	2.37
T2	-1.82	2.46	T8	not present	1.08
T3	-1.87	3.71	T9	not present	3.72
T4	-0.96	1.22	T10	not present	3.01
T5	-1.59	0.78	T11	not present	4.24
T6	-2.73	3.19			

Scanning the values in this table reveals that Initiation behaviors disappeared after 40 minutes (starting at interval T8) and did not reappear. Figure 9 shows the overall pattern of changes in the level of engagement required to transition between categories of behavior over the hour of observed activity.

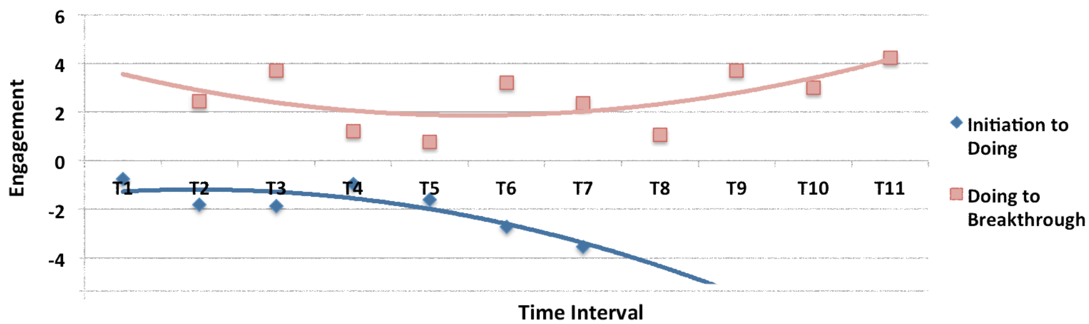

Figure 9. Amount of engagement required to transition between behaviors over time during bioblitz specimen sorting.

As was true for the DNA barcode transition thresholds in Figure 6, transitions from Initiation to Doing became easier over time. In fact, the level of engagement required to begin Doing decreases even more quickly than for the work with DNA barcodes and computers This might be due to the heterogeneous nature of the group: the novices who were not sure about transitioning to Doing left, while the others eventually learned enough by watching or through help by others to settle down to work.

Figure 6 and Figure 9 also show the same pattern of changes in engagement required for transition to Breakthrough behaviors; the amount of engagement required decreased over the first half of the observation period and increased toward the end. As was true for the observations during the DNA barcode work, some of the behaviors categorized as Breakthrough during the bioblitz sorting involved more expert participants providing help to others. We wondered what would happen if we once again re-categorized such behaviors as instances of Doing. Would the line once again flatten out? Figure 10 summarizes the result.

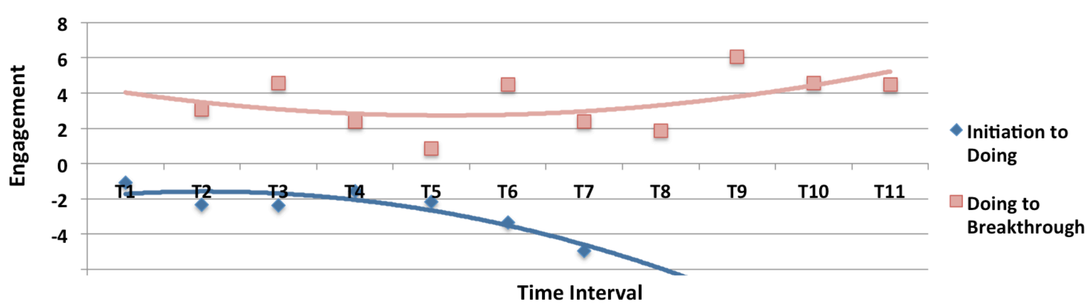

Figure 10. Change in engagement required for Breakthrough after adjusting observation framework: bioblitz sorting.

Adjusting the observation framework had the effect of flattening the curve for the amount of engagement required for Breakthrough to some extent, but the upward trend is still evident. Once again, one can generate conjectures about what is happening here. Perhaps as less experienced participants either settled down to work or left, there was more doing and less sharing. Or, perhaps we should think about other adjustments to the observation framework. The point of this example, in terms of the present paper, is that using the Rasch partial credit model to create a measurement scale for engagement assists in stimulating such thinking.

Discussion, Contribution, Limitations

As noted its opening paragraphs, the present paper is an exploration of a methodological problem: we have been unsatisfied with our efforts to develop and refine participant observation frameworks when the outputs from those frameworks are conceived only as counts of different categories of behavior. We were familiar with use of Rasch models from our work in formal education settings and so wondered whether it might be useful to apply a Rasch model to assist in refining an instrument that uses observation data to make estimates of participant engagement.

Using two small data sets and an observation framework that is still in the early stages of development, this paper demonstrates that Rasch analysis using a partial credit model produces results that are generally consistent with simpler, less quantitatively rich frequency count data. This is an important result. As we set out on this exploration, it was not clear to us that the Rasch model would fit the kinds of observation data that we were collecting. The fact that the Rasch model did fit reasonably well with these small datasets is encouraging. Rasch analysis is increasingly useful and reliable as the number of participants becomes larger, encompassing hundreds of participants over multiple observation sessions. The fact that we are seeing coherent and potentially useful results in these pilot-study datasets suggest that using the Rasch partial credit model will be useful with larger datasets.

Contribution to Citizen Science Research

In this paper we used Rasch partial credit models to assist us in gaining insight into the performance of an observation protocol focused on participant engagement. The approach outlined here is not limited to measures of engagement. Research into the practice of citizen science might rely on observation data in investigating a variety of important theoretical constructs such as persistence, curiosity, capability, and others. In exploring each of these theoretical constructs, having a way to transform counts of observations into an interval scale that represents the amount or degree of the construct has the potential to assist in instrument design and in refining the definition of the construct.

As we noted in the first paragraphs of this paper, researchers or evaluators who are designing observation frameworks to measure engagement, persistence, curiosity, capability, or something else must address two questions: (1) "Which behaviors should be recorded?" and (2) "What values should be assigned to these behaviors, corresponding to more or less of the construct that we are measuring?"

These two questions—one focused on matters of definition and the other focused on scaling—are connected. The exploratory work reported here shows how addressing the scaling question can inform the definitional question. Attention to scaling can assist in framing and gathering data to answer questions such as:

- Is the instrument measuring one thing or several things?
- Are the behaviors that are grouped together really of a piece, or should they be categorized in some other way?
- Are the assumptions that we make about categories of behaviors that we believe are associated with the construct supported by the data? For example, is category 1 consistently less demanding than category 2?

Limitations

As already noted, this paper describes exploration of a methodological problem that arose in the course of pilot work focused on broader questions. It reflects the limitations that are often inherent in exploratory and pilot work. In this case those limitations include use of small samples of participants and the use of tools, such as the engagement observation framework, that are still under development.

Because sample sizes are small, the work presented here should not be used to make claims about the DNA barcoding and bioblitz activities that we used in these analyses. This paper focuses on exploration of a measurement problem, rather than on making generalizations about the programs from which we drew the data used for exploration.

Suggestions For Future Work

We propose two different areas for future exploration and development on the basis of the work presented here. The first focuses on designing observation frameworks. The second extends more broadly to collaborative exploration within the community of researchers and evaluators seeking to understand the potential and the effects of citizen science programs.

Observation Framework Design

The exploration presented here suggests potentially fruitful additional research that might be pursued to improve the sensitivity of the observation framework and to tie it more firmly to a clear conception of engagement. The additional research would involve increasing the number of categories of behaviors and narrowing their focus. The observation framework presented here combines a number of different behaviors into the three broad categories of Initiation, Doing, and Breakthrough. As described above and illustrated in Figure 1 and Figure 2, these categories represent *a priori* conceptions of engagement.

Given the increased confidence emerging from this exploratory work that the Rasch partial credit model can be usefully applied to observation data, it would be useful to experiment with moving from grouping behaviors into categories toward making direct use of different observed behaviors (for example, sharing and picture taking) in the analysis. This would result in estimates of the engagement associated with each behavior, rather than an estimate of engagement for a category that is an amalgam of different behaviors. Doing this would require a great many more observations than we had available during this pilot work, since it would take many more observations to provide a rich picture of nine or ten individual behaviors, as opposed to just three general categories that group behaviors together. It is possible that such a more fine-grained investigation would lead back to some kind of grouping of behaviors, but it would have the advantage of grouping behaviors on the basis of what was learned from observation data, rather than grouping them on an *a priori* basis.

Broader Collaboration

The community of researchers and evaluators who study the design and outcomes of citizen science projects is already engaged in a number of collaborative endeavors that enable different researchers and evaluators to share and build on each other's work. For example, the DEVISE project has assembled and developed a suite of instruments to measure constructs such as interest, scientific self-efficacy, and skills (Phillips, et al., 2014), with the aim of stimulating use of measures that might be compared across studies.

We see the potential for a similar, or perhaps a related effort that could grow out of the exploratory work presented here. Such collaborative work could unfold in two stages.

In the first stage different researchers could collaborate in pooling observation data to create better observation frameworks capable of providing interval scale measures of constructs such as engagement, persistence, and so on. As noted above, the power and utility of the measurement tools explored in this paper increases substantially as they are applied to larger numbers of participants. Specifically, with access to many more observations of citizen scientists, it would be possible to explore observation frameworks that make the kinds of more nuanced distinctions described in the preceding section. Developing this kind of more powerful, reliable, and valid measure of different constructs would proceed much more quickly if the citizen science research community pooled its observation data.

During this first, design and construction stage the goal would be to produce a small number of frameworks aimed at use with different participant populations (for example, expert citizen scientists and more casual citizen scientists) or in different settings (for example, sample collection in the field as opposed to laboratory work). Working iteratively over time and using data from many studies and using analytical insights provided through Rasch modeling, these observation frameworks might develop into a suite of measurement instruments with well-understood properties and applications.

The second stage would focus on using, as opposed to refining, the observation frameworks emerging from the first stage of work. The aim of this effort would be similar to a second aim of the current DEVISE project: create a set of tools that facilitates comparison and learning across different programs, contexts, and participant populations. Having access to observation frameworks that resulted in interval scale measures, as opposed to frequency counts, would assist in making statistical inferences about gains, changes, and differences across studies.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. DRL-1223210. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Barriault, C. (2014). Assessing Visitor Learning in Zoos and Aquaria: A Revised Framework. Presented at the NARST Annual International Conference, Pittsburgh, PA, March 30-April 2, 2014.

Barriault, C., & Pearson, D. (2010). Assessing Exhibits for Learning in Science Centers: A Practical Tool. *Visitor Studies*, *13*(1), 90–106. doi:10.1080/10645571003618824

Bond, T. G., & Fox, C. M. (2012). *Applying the Rasch Model* (2nd ed.). New York: Routledge.

Horizon Research. (2000). Inside the Classroom Observation and Analytic Protocol. Chapel Hill, NC: Horizon Research Inc. Retrieved from http://www.horizon-research.com/inside-the-classroom-observation-and-analytic-protocol/

Mair, P., Hatzinger, R., Maier, M. J., & Rusch, T. (2015). Package “eRm.” CRAN. Retrieved from http://cran.r-project.org/web/packages/eRm/eRm.pdf

Palmquist, S. D., & Crowley, K. (2007). Studying dinosaur learning on an island of expertise. In R. Goldman, R. Pea, B. Barron, & S. J. Derry, *Video Research in the Learning Sciences* (pp. 271–286). Mahway, NJ: Routledge.

Phillips, T. B., Ferguson, M., Minarcheck, M., Porticella, N., & Bonney, R. (2014). *User’s Guide for Evaluating Learning Outcomes in Citizen Science*. Ithaca, NY: Cornell Lab of Ornithology.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press.