Investigating the Cascading, Long Term Effects of Informal Science Education Experiences Report

John H. Falk , Judith Koke , C. Aaron Price , Scott Pattison

Suggested citation: Falk, J. H, Koke, J., Price, C. A. & Pattison, S. (2018). Investigating the Cascading, Long Term Effects of Informal Science Education Experiences Report. Beaverton, Oregon: Institute for Learning Innovation.

Introduction

Science educators invest their time, effort and expertise in the effort to "make science enjoyable and interesting" and "inspire a general interest in and engagement with science." The consensus purpose for science education experiences are the long-term transformation of the learner; an individual sufficiently engaged that s/he will have the interest and tools necessary to pursue a "cascade" of experiences subsequent to the initial educational event. Support for this conclusion comes from a large-scale survey of both informal and formal science educators in the United Kingdom where 90% of science educators from across more than a dozen sectors, including museums, science centers, zoos and aquariums, print and broadcast media, outdoor facilities, libraries and schools, rated the above two educational goals as their highest priorities (Falk, et al., 2015). A laudable purpose indeed, and many dedicated professionals work to that end, but the question remains: How can the field measure whether or not any particular educational experience sets off long-term, cascades of additional learning experiences?

Science learning is neither linear nor easily isolated in time and space.

A number of major challenges limit valid and reliable documentation of the long-term effects of any science education experience. Two are particularly vexing, the complex and cumulative nature of science learning and the inherent limitations of current research methods and tools.

The Complexity of Learning

It was once assumed that learning, and science learning in particular, was a straightforward, linear process that primarily occurred through directed instruction, i.e., the absorption-transmission model (cf., Bransford, Brown & Cocking, 2000; Roschelle, 1995). However according to a recent OECD publication (Dumont, Istance & Benavides, 2012), the dominant view of learning, today, is a socio-constructivist view, in which "learning is understood to be importantly shaped by the context in which it is situated and is actively constructed through social negotiation with others" (p.3). From this perspective, any particular learning experience, whether it takes place within a classroom or a science museum, is almost certainly influenced by a host of other learning experiences that occurred previously in a person's life. Thus, the ultimate outcome or effect of a particular learning event is likely to be only a partial consequence of that specific event. A full accounting of even short-term effects would require knowing something about each learner's unique learning history prior to the event, and then only by viewing these trajectories in the aggregate could some understanding of the overall outcomes/effects of that event be inferred (cf., Falk, 2018).

In other words, learning is rarely, if ever, instantaneous (Bransford, Brown & Cocking, 2000). Individuals develop an understanding of and appreciation for scientific topics through an ongoing accumulation of experiences and understandings derived from multiple sources (e.g., Anderson, Lucas, Ginns, & Dierking, 2000; Barron, 2006; Bathgate, Schunn & Corenti, 2013; Bell, et al., 2013; Bransford, Brown, & Cocking, 2000; Falk & Needham, 2013; Ito, et al., 2013; Lemke, 1999; NRC, 2015). For example, an individual's understanding of the physics of flight might represent the cumulative experiences of completing a classroom assignment on Bernouli's principle, reading a book on the Wright brothers, visiting a science center exhibit on lift and drag, and watching a television program on birds. For the individual, all of these experiences are combined, often seamlessly, as they construct a personal understanding of flight; no one source is sufficient to create understanding, nor one single institution solely responsible. In the above scenario, when did this individual learn about flight and what experiences most contributed to learning? And how could one specifically identify and attribute the pieces learned while at, for example, the science center as opposed to the pieces learned in school, reading, or television? In summary, science learning is neither linear nor easily isolated in time and space.

Methodological Issues

This leads to the second major set of challenges in measuring the cascading events that constitute science learning—methodological challenges. Historically, the vast majority of efforts designed to measure the consequences of a science education event were limited in both duration and scope. The most common measures, in both informal and formal contexts, utilized some kind of pre-post design which measured changes in understanding, attitude,

etc., based on responses to some kind of test administered immediately preceding and immediately following a particular educational event. The assumptions of this approach being that: 1) within this short timeframe changes in understanding, interest, etc. should emerge; and 2) any changes that do occur, are directly attributable to the educational interventions of that event. As should be clear from the above review, these assumptions may or may not be actually true. Changes in the mental structures, i.e., learning, often take time to emerge and in the absence of suitable preexisting structures and scaffolds, as well as the presence or absence of subsequent reinforcing events, may not persist (Eaglemen, 2015).

In response to this first issue, a number of investigators, including particularly many within the informal/free-choice realm, have attempted to lengthen the timeline of assessment by weeks, months and even years (e.g., Adelman, Falk & James, 2000; Bell, et al., 2013; Falk & Dierking, 2014; Falk, et al., 2004; Flagg, 2005; Fraser, et al., 2012; Peterman, Pressman & Goodman, 2007). Naturally, the longer the timeframe, the greater the challenge in maintaining contact with individuals and the potential for introducing other types of biases into the data. For example, panel/longitudinal designs are notorious for becoming less representative over time as the population changes and as panel members drop out (Groves, 1989; Taplan, 2005). Longitudinal designs may also be prone to certain forms of measurement error, such as "conditioning" and "seam" bias (cf., Groves, 1989; Lavrakas, 2008). Another problem is that virtually all of these longitudinal studies, including those cited above, utilized in whole or in part self-report data; an approach long viewed with skepticism with the social science community (cf., Baer, Renaldo & Berry, 2003). Although a number of studies from various disciplines have established that self-report data, though not perfect, are actually reasonable surrogates for more direct measures, especially when using survey data (Chan, 2009; Gonyea, 2005; Vaske, 2008), finding alternative or at least additional measures to support the validity of changes would seem important.

Equally problematic, and potentially even more intractable, has been the challenge of attribution. Given the inherently incremental and distributed nature of science learning, how can one be certain that any observable changes in an individual's knowledge, interest or behavior are actually attributable to the experience under study? Very few studies have seriously dealt with this issue.

Responding to the Challenge

A group of 12 researchers gathered at MSI to discuss issues surrounding the critical but challenging area of how to measure the long-term effects or impacts of ISE experiences, chaired by Dr. John Falk, Oregon State University and Institute for Learning Innovation, and hosted by Aaron Price, Museum of Science & Industry, Chicago, (MSI) on July 18, 2018. All the invitees had expertise and experience in this area, resulting in a rich conversation based equally in theory and practice. The goal of the discussion was to build on the collective experience of the group to identify and address key challenges in this area of research and to propose potential solutions for advancing the field. This whitepaper provides a summary of those deliberations.

The day was divided into three time blocks. The early morning was devoted to large group discussions about long-term learning research and its inherent issues and challenges, but the bulk of the day was spent in three smaller working and writing sub-groups; each group focused on one of three key issues:

1. Timelines and slopes (how to determine when to collect "long-term" measures; what constitutes long-term; what are the "slopes" of effects?)

2. Attribution (effects vs. impacts; how to determine causal relationships)

3. Accommodating changes in context (e.g., how to deal with the impact of major shifts in culture and society that occur in the midst of long-term measures) and whose concept or definition of "success" should one measure (the public's, ISE staff's, ISE institutions?)

A designated facilitator and writer led each sub-group. After both a morning and afternoon of intensive conversations, the leaders shared a summary of the conversations with the larger group for one final whole-group, reflection session.

Each of the small group facilitators compiled a written summary of their group's discussion that was then circulated to their group members for comment and approval. Every effort was made to ensure that these summaries equitably represented the contributions of all group members. The sections below are the products of these efforts.

Question 1: Timelines and Slopes

Aaron Price, John Falk, Gail Jones

Measuring change over time is a fundamental challenge to almost all sciences (Duncan & Duncan, 2009). In education and many other social sciences, intervention goals are often focused on achieving long term and/or permanent outcomes, although most measurements of success are usually very short-term. In informal/free-choice settings, this dual reality is equally true. Interventions in such settings are often of relatively short duration, creating a greater dilemma than in many other educational contexts. For example, an after-school program may work with children for a few hours a week over the course of a year or two. And a museum exhibit may gather the attention of a guest for just a few minutes. What kind of long-term outcomes can one expect from such short interventions? Happily, an abundance of research has shown that even these very brief experiences can have very positive effects (see reviews by Anderson, Storksdieck & Spock, 2009; Falk & Dierking 2018; NRC, 2009). However, despite considerable research and support for the existence of some kinds of positive long-term effects for informal/free-choice learning experiences, the specific nature of these long-term effects is less well known. In particular, questions remain about the general duration (e.g., do effects last days, weeks, months, years?) and character (e.g., do effects wax and wane or do they just happen?) of such learning. The challenge to both researchers and education programmers is how to best use their limited resources to construct the most rigorous methodology or intervention, yielding the best results. In other words, what is lacking is a robust model of change. Having such a model or set of best practices would go a long way in helping researchers and practitioners design long-term studies/experiences within informal/free-choice learning settings.

Education is not the only domain to have been concerned with this topic. We could learn from the many other fields that are interested in long-term outcomes. Medicine comes to mind first with its emphasis on long-term impact of both interventions and side effects. When it comes to drug interventions, long term effects are often described using dose-response relationships (Farinde, 2017). The relationships are described using key factors such as potency (of the intervention/drug), slope (change of effect over time) and maximal efficacy (time and significance of peak impact). In the social sciences, analysis of long-term change is common through the use of latent growth models (LGM) (McArdle & Epstein, 1987; Duncan & Duncan, 2009; Isiordia & Ferrer, 2018). Long-term effects are also used as a component of structural equation modeling. These flexible statistical models are used to describe and predict growth and change over time in a manner that allows the researcher to include many of the complexities of social science data (such as multiple contextual variables or correlated responses between participants).

One way to visualize long-term outcomes are with slopes of effect (hereafter: SoE). Similar to growth curves and dose-response relationships, the slopes show change over time from a baseline level and with an intervention moment. However, they do not assume positive, linear or sustained growth. Typically, the dependent variable would be displayed on the vertical axis and time on the horizontal. Considering that human nature loves to work in logarithms (Dehaene, Izard, Spelke & Pica, 2008),

one may want to begin with an inverse logarithmic function as the basis for a typical SoE. But the more complex the model becomes, the more degrees of freedom are needed to characterize it. Figures 1–4 show four hypothetical learning-based SoEs.

Implications and Recommendations:

Researchers designing long-term studies may want to draw hypothetical SoEs as a thought experiment to help them determine when to measure outcomes and, combined with power analysis, how much data to collect at each point along the time continuum. As one can imagine, the more complicated the slope the more sampling points across time are needed to fully measure it and, as always, the less steep the slope, the more data required.

Figure 1: Forgetfulness

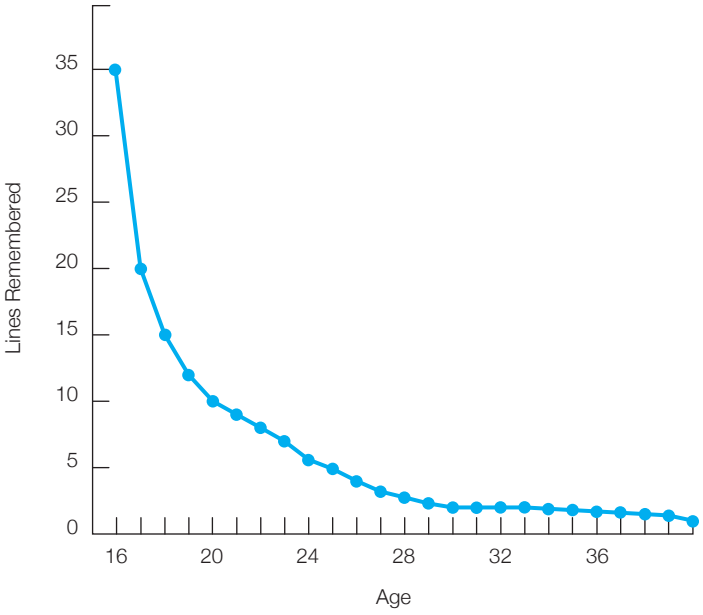

Figure 1: A hypothetical inverse logarithmic decline. In this case, a student memorized the lines to Lewis Carroll's Jabberwocky for a class project, then forgot them over time.

Figure 2: Hot Stoves Hurt

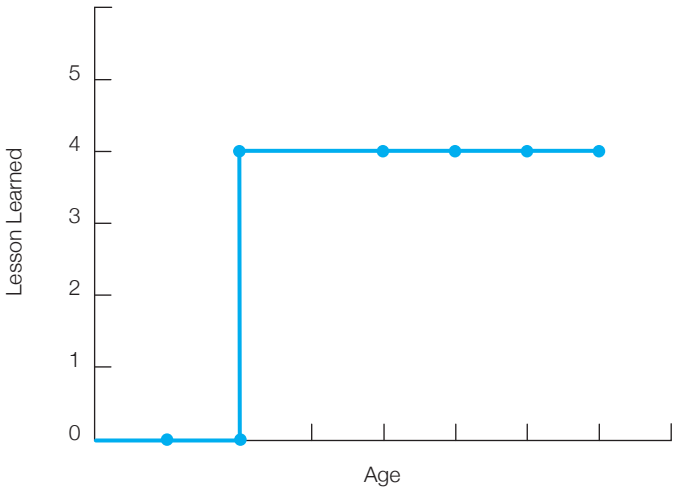

Figure 2: A child touches a hot stove and learns a valuable lesson which they never forget.

Figure 3: Short Museum Visits

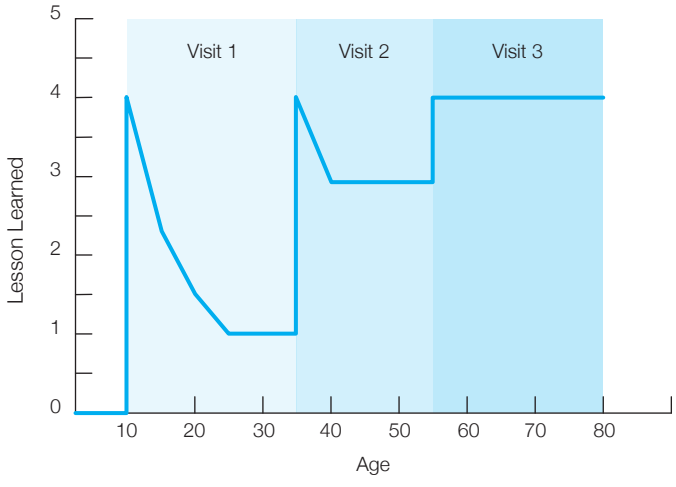

Figure 3: Someone visits a museum as a child and learns about butterflies in an exhibit. They slowly forget what they learn only to have it reinforced when they visit again with their own children. This time, they recall more than before, but still forget some content until they visit again—this time with grandchildren. At this point, they remember the specific information.

Figure 4: Long Term Engagements

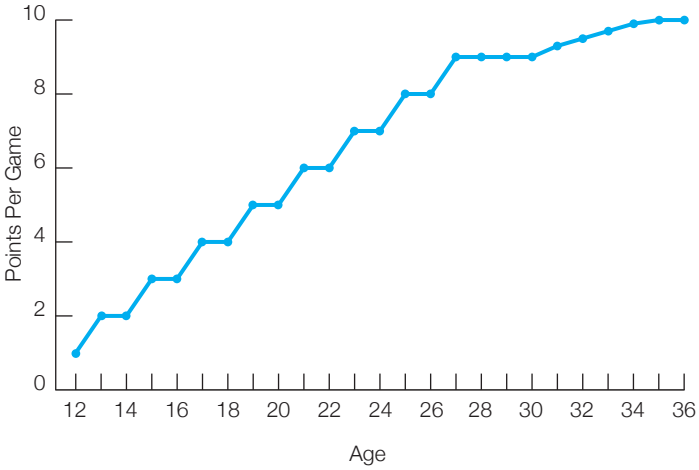

Figure 4: A child starts playing basketball at school, six months per year annually. Every year, they get a little better during the season. Then, after high school they only play recreationally. They still improve, but at a slower pace.

Question 2: Attribution

Scott Pattison, Lynn Dierking, Robert Tai, Jim Kisiel

Serious issues surround the question of attribution of causal relationships related to measuring the long-term impact and influence of informal/free-choice learning experiences. In other words, to what extent can measured outcomes and effects that are found to relate to specific informal/free-choice learning experiences be claimed to be directly or indirectly caused by these experiences? This issue is central to research, policy, and practice in the field. A long-standing question for researchers studying learning inside and outside of school has been the extent and degree to which informal/free-choice learning experiences and institutions contribute to learning outcomes for children and adults, as opposed to these outcomes being a result of selection bias or other confounding factors (e.g., Falk & Dierking, 1997; Falk, Dierking, et al., 2016; Falk & Needham, 2013; Falk, Pattison, Meier, Bibas, & Livingston, 2018; Tai, Liu, Maltese, & Fan, 2006). The answer to this question has profound implications for not only educators working in these spaces, who have advocated for the importance of informal/free-choice learning (National Research Council, 2009), but also policy makers and funders, who historically have been skeptical about investing significantly in education initiatives outside the formal school system (Falk & Dierking, 2010).

In our discussion, three themes emerged that were further supported and informed by the pre-workshop readings and other resources shared by participants:

1) Studying questions of attribution and causality is important for the field.

2) Attribution and causality are more complex than assumed by many traditional models, such as pre/post experimental studies.

3) Testing and supporting claims of attribution and causality is an ongoing process, beyond a single study.

The group also identified communication during and after the research process as a cross cutting topic that was relevant to each of the themes. Finally, the group discussed implications and recommendations related to this emergent understanding of attribution for understanding the long-term impact of informal/free-choice learning experiences.

The Importance of Attribution

Many scholars have highlighted the challenges of making definitive claims about the causal impacts of informal/free-choice learning experiences on participant outcomes or of assessing how the contributions of these experiences relate to the ongoing, cumulative nature of learning across a person's life, inside and outside of school (Falk, Dierking, & Foutz, 2007; National Research Council, 2009; Pattison & Shagott, 2015). Some have even suggested that the focus on simple causality and attribution is misguided or antithetical to the idiosyncratic, free-choice nature of informal learning experiences and institutions (cf., Falk, et al., 2016a).

The group was unanimous in agreeing that it is critical for researchers to study these causal relationships and test hypotheses about the direct and indirect impacts of informal/free-choice learning experiences on long-term participant outcomes. Researchers in this field have prioritized research-practice connections and often work directly with and alongside educators and policy makers to support and enhance learning outside of school. These educators and policy makers are, in turn, focused on using research to make decisions about what types of experiences and practices lead to positive learning effects over the long term and how funding and other resources can best be invested to support learning. This includes decisions about the relative focus on traditional education institutions, such as schools, and emerging institutions and systems outside of school or across formal and informal education settings. At the core, these are questions about causal relationships, even if the causal mechanisms and processes are not always simple, direct, or immediate. Thus, the group agreed that it is not enough for the field to assess correlational relationships or provide anecdotal evidence of impact (although those types of studies can provide important contributions to causal questions). Instead, researchers must work with educators and institutions to find innovative ways to test assertions about the causal relationship between informal/free-choice learning experiences and the long-term effect on participants.

The Complexity of Attribution

Although the group agreed that focusing on causal relationships is important, much of the discussion focused on how approaches to studying, describing, and communicating these relationships need to become much more nuanced and sophisticated in order to account for the complexity of attribution and causality related to long-term learning outcomes. Traditional perspectives on causality and attribution focus on the classic formulation of establishing temporal order, measuring correlation, and accounting for confounding factors or alternative explanations (Campbell & Stanley, 1967; Fu, Kannan, Shavelson, Peterson, & Kurpius, 2016; Shadish, Cook, & Campbell, 2001). However, many other perspectives on causality and approaches to studying causal relationships have emerged that recognize the complexity of learning in the real world (Gates & Dyson, 2017; Lemke, Lecusay, Cole, & Michalchik, 2015). For example, some perspectives emphasize feedback loops and emergent properties within complex systems, rather than linear relationships, or highlight the importance of participant narratives of causal chains. In the field of evaluation, there has been a growing focus on contribution rather than attribution, recognizing that the impact of any single program or initiative will be influenced by the variety of other experiences in a person's life, before, during, and after the program takes place (Gates & Dyson, 2017). And research in informal/free-choice learning settings has consistently highlighted the important influence of what individual participants bring with them to the experience and what happens to them afterwards. Even within a traditional framework of thinking about causality, causal relationships almost always represent averages or probabilities, rather than universals, certainties, or inevitabilities. Modern statistical techniques allow researchers to model and test complex causal chains, multiple contributing factors, and mediating and moderating relationships, which often reveal the nuances underlying simple relationships between experience and outcome.

Overall, the group recommended that researchers seek to study and communicate more nuanced hypotheses about causality that reflect the situated and contingent nature of these relationships. For example: These types of experiences will likely lead to these types of outcomes for these groups of participants in these particular circumstances. As discussed more below, this formulation of a causal relationship makes clear the ongoing work, beyond a single study, that is needed to provide evidence for how the relationship generalizes to different contexts and different participants and the contextual factors and contingencies that influence the strength or probability of that relationship. Similarly, the group discussed the interplay between internal validity, or the strength of the causal relationship between cause and effect as conceptualized in a traditional experimental study, and other related issues such as external validity (i.e., is the causal claim relevant to contexts beyond the research study) and generalizability (i.e., does the causal relationship hold true for other participants, contexts, programs, etc.). Although clarifying the limitations of such studies is critical for the field, helping decision-makers avoid blind acceptance or overly critical rejection of findings becomes an equally important role for those researchers exploring long-term effects.

The Ongoing Study of Attribution

Because causal relationships are complex, situated, and contingent, the group discussed the importance of moving beyond a focus on single studies and instead developing bodies of research that cumulatively test and explore the limits of hypothesized causal relationships. A statement of a causal relationship is in essence an argument about a hypothesized connection between one or more causal factors and one or more outcome measures. Any single study can only provide partial and imperfect evidence to support, or contradict, that causal argument. Researchers, therefore, must clearly understand and communicate the strengths and limitations of the evidence within a given study and strive to test and explore those strengths and limitations with other study findings, different data sets, and further investigations. The group agreed that it is the responsibility of the researcher to be: transparent about methods and their connection to evidence and claims; frame claims appropriately given the level of evidence and limitations of a particular study; describe the study context in which those claims are situated; use language about causal relationships carefully and appropriately; and help outline for the field the next steps needed to continue investigation of the underlying causes and attributions motivating the study.

Similarly, it is the responsibility of the field and consumers of research to ask critical questions and foster a professional culture of respectful and productive debate about evidence, claims, and methods. Researchers can model these expectations by inviting and responding productively to study questions and critiques. Ideally, the group discussed how this ongoing

process of testing causal attribution should be informed by direct studies of causal relationships as well as explorations of the processes and mechanisms underlying those relationships, which can shed light on or help nuance the understanding of the relationships themselves (Shadish et al., 2001). In this process, quantitative, qualitative, and mixed-method approaches are all important for contributing to a deeper understanding of questions about cause and attribution.

The Communication of Attribution

During the group discussion, communication emerged as a cross-cutting topic relevant to all three of the themes described above. This communication involved how researchers describe and share study methods and findings, how these messages are tailored to different audiences, and how researchers collaborate with educators and policy makers.

As noted, the group extensively discussed the responsibility of researchers in carefully articulating their causal claims, study methods, level of evidence provided by the data, limitations and alternative explanations, and questions for future work. However, this type of technical language is often not appropriate or understandable by non-researcher audiences, such as educators and policy makers. The group noted that the allure of simple causal arguments and immediately actionable findings is strong for many stakeholders, including funders, which can make it difficult for researchers to situate the limits of a single study or set of data appropriately. Addressing this issue involves not only finding new approaches to communication and ways of collaborating with educators and policy makers, but also new efforts to develop shared understandings about the scientific process. All stakeholders interested in the effects of informal/free-choice learning need to understand that the development of knowledge through scientific research is incremental, ongoing, and imperfect; such is the nature of science. Understanding the limitations of how science and research work (i.e., claims are based only on evidence, evidence may be interpreted in different ways, and absolute certainty is impossible due to the fact that new information may require revisions to claims) is important when making sense of these investigations and the reported outcomes. The strengths, weaknesses and limitations of studies that examine these complex interactions must be made clear to stakeholders, who in turn must examine such work with a critical yet informed perspective.

In addition to communicating the limitations of particular studies or findings, the group agreed that progress on understanding causal relationships related to the long-term effects of informal/free-choice learning experiences are only possible through ongoing communication and collaboration between researchers and practitioners. Educators and practitioners working directly with programs and participants can help researchers understand the complexities of those experiences, what causal relationships and pathways are worth studying, and how those might be captured through research methods, measures, and analyses. Similarly, researchers can help practitioners develop program models and theories of change that account for the complexity of long-term impacts and can provide evidence to help guide decisions about the educational strategies and programs that are likely to lead to those impacts. Both researchers and educators can work together to pursue approaches to supporting learning that recognize rather than ignore the individual, situated, and contingent nature of informal/free-choice learning, such as allowing for personalization and creating ongoing support beyond a particular program or experience.

Challenges and Opportunities for Studying Attribution

In addition to the ideas above, the group discussed other challenges and opportunities inherent to studying attribution and causality related to the long-term effects of informal/free-choice learning experiences:

- Designing for complexity including methods and analytic approaches that capture the complexity of factors influencing long-term effects and creating broader program initiatives that influence or strategically align multiple experiences for greater impact.

- Understanding variation across informal/free-choice learning contexts related to attribution, such as different levels and types of possible outcomes from a single museum visit or a year-long afterschool program.

- Training and supporting researchers and evaluators in the field to use a variety of tools and perspectives, including approaches from other fields (e.g., Baer, 1988; Fivush, McDermott Sales, Goldberg, Bahrick, & Parker, 2004; Howell, 2014; Peterson, 2002), to pursue questions about causality and attribution.

- Balancing the roles of advocate and researcher, especially related to the honest and transparent communication of study limitations, the nuanced and careful communication of causal claims, and the ongoing debate

about the relative value of informal/free-choice and formal learning experiences.

• Creating mechanisms to support ongoing dialogue about study findings, methods, and causal claims that advance the field's understanding of attribution and approaches to studying long-term impacts.

• Finding sufficient time and resources to conduct this type of rigorous research and implement programmatic and educational strategies that reflect complex ideas about attribution and contribution.

Implications and Recommendations

How can informal/free-choice learning scholars and institutions grapple with the complexities and challenges described above and make headway in testing the effects of out-of-school experiences on long-term learning experiences? Affirming that it is no longer enough for the community to dismiss causality as an impossible goal, the group discussed three strategies for moving forward:

1) Focus and commit to values and effects—As discussed more extensively in other sections of the workshop, a key challenge to measuring long-term effects is defining and operationalizing what effects or outcomes should be prioritized. By focusing and articulating priorities, this group argued that institutions and programs can greatly increase the likelihood that educational strategies will be designed to effectively achieve outcomes and that the assessment of those outcomes will be aligned with the effects that the program or institution is best positioned to support.

2) Find methods for accounting for the complexity of attribution—Because of the variety of factors and complexities potentially influencing how and whether a particular learning experience will lead to long-term learning outcomes, the group agreed that it is important for researchers and educators to collaborate to find creative and innovate methods for accounting for this complexity when developing and testing models of long-term effect. For example, understanding the initial motivations of visitors to a museum can reveal how the outcomes of those experiences differ by participant group (Falk, 2009; Falk & Storksdieck, 2005). This may be especially important when the experience or program is relatively brief and the long-term effects will likely vary greatly across individuals based on prior and subsequent experiences. One way of thinking about this strategy is measuring and accounting for the "noise" within a learning system in order to better identify the "signal" showing the causal relationships or pathways between an informal/free-choice learning experience and long-term learning outcomes.

3) Design initiatives and partnerships to influence the complexity of attribution—A different approach to tackling the complexity of attribution is to design educational systems and partnerships that have a broader influence on the many learning contexts and experiences across a person's life and thus exert greater, more synergistic influence on long-term learning effects (e.g., Falk, et al., 2016b; Pattison et al., 2017). This is aligned with the growing emphasis within the field to think about and support the learning ecologies of children and adults and help learners build on experiences across contexts and institutions (National Research Council, 2009, 2015). Beyond simply account for the "noise" within the system, this approach focuses on designing support structures so that the "noise" becomes part of the "signal" that increases the likelihood of achieving long-term impacts that can be attributed to specific educational programs and experiences.

Question 3: Accommodating Context

Judith Koke, Kate Livingston, Ali Mroczkowski, Rabiah Mayas

The purpose of researching longer-term effects is usually to attempt to understand intervention impacts and outcomes that take time to take place. Rather than measuring immediate changes in knowledge, interest or attitude, a longer-term investigation usually focuses on the persistence of these changes, and on any resulting consequences, e.g., behaviors, of these changes.

In shorter term program evaluation/research, the metrics of outcomes are very program-specific, which often excludes findings related to organizational mission, community impacts, or individual outcomes not defined in advance. The move to outcome evaluation in the field, while a significant step forward, has significant shortcomings in that they:

• Make assumptions about accessibility and equity, e.g., assuming that everyone has the same access to an experience and failing to acknowledge issues of equity in determining outcomes (For example: an equitable assessment of graduation rates in STEM would require acknowledging that SES and parental support differences are as important in determining outcomes as are what happens within the K-12 grade system.)

• Require that "we" inside the field disproportionately define what is important to measure, usually in a prescriptive manner, rather than in collaborative way that recognizes individual, cultural and community defined needs and priorities.

• Provide a barrier to performing open ended research focused on understanding what's important to measure.

• Often skew towards the funders' interests, which may or may not necessarily be fully aligned with what is deemed most valuable or interesting by the program organizers or the program's participants.

Given these shortcomings, considerably more thought needs to be given to identifying and capturing not only the intended goals of the institution but equally the needs and goals of the intended audiences, including and specifically the likelihood that due to cultural and experiential differences, different audience constituencies may have different needs and goals.

Accordingly, it is important to consider that:

• Participant reflections on an experience overall may not be positive, but there can still be positive gains/effects/perceived moments of value;

• Some program participants can be doing meaningful things that are valuable to them, but the program/intervention/research is disrupting what they see as the best use of their time;

• In some investigations, and with some outcomes (e.g., behavior, interest), it will be important to make room for/acknowledge that cultural differences will likely effect outcomes;

• The scientific enterprise has its own set of discrete norms and cultures which directly influence how research is done even though there may situations and circumstances where it may not always be appropriate for these norms and cultures to be privileged;

• It is important that the field begin moving towards more individualized approaches to measuring impacts, pathways, and learning; all while maintaining high standards of research validity and reliability;

• Project timelines impose a tremendous constraint on validly measuring long-term effects; particularly given the significant time it takes to build trusted relationships necessary for successful research within many, particularly underserved communities;

• Research models might want to include within measures of effect how educational experiences contribute to the wellness of a community/family (as defined by the learners, not institutions);

• Research designs/approaches need to include measures that capture multiple levels of effect—effects at the level of the individual, the program and the community (cf., NRC, 2015);

• Some long-term metrics need to be open-ended so that learners themselves can have agency in defining where and how impacts have occurred in their lives; and finally

• The field still strongly privileges quantitative research over qualitative work, yet investigating longer-term, culturally-specific individual effects are likely to be most readily discernable through data-rich, qualitative approaches.

Implications and Recommendations:

The group makes the following suggestions for how to positively move the field forward. There is a need for the field to:

1. Place greater emphasis on reflexive practice—identifying specific moments in the process to review and modify the research/evaluation process in response to investigator learning (cf., Michael Quinn Patton (2016) Developmental Evaluation process).

2. Practice greater cultural humility. As our field begins to increase attention to and skill with culturally sensitive research we need to move away from prescriptive, linear approaches and open-up the research process to be more inclusive of its subjects. More qualitative, open explorations of the journey and the definition of benefits from the participants' perspectives should be encouraged. We must consider what an engaged, inclusive meaningful role for the participants could be. Can we involve subjects in shaping the questions and/or in responding to the interpretation of data?

3. Demonstrate greater professional humility: Considering the increasing perception that conducting long-term investigations is a necessity rather than a nicety, the research questions inherent in valid and reliable long-term investigations must drive the methods—not the reverse.

However, we perceive significant barriers or challenges to moving informal science education research in this direction, in particular:

1. Quantitative research is more highly valued than qualitative research—yet this work requires the inclusion of qualitative approaches.

2. More inclusive approaches are very time consuming—and hence more expensive.

3. Is our field actually open to change, or will practitioners, funders and even some researchers find it threatening?

Conclusion

This highly productive meeting clearly identified some key challenges and issues inherent in investigating the long-term effects of informal science education experiences. As summarized below, it also generated some basic recommendations for both practice and policy.

1. Researchers designing long-term studies should consider illustrating the presumed learning process through a hypothetical Slope of Effect. This demonstration will work to help determine when to measure outcomes and, combined with power analysis, will help to clarify how much data to collect at each point along the time continuum.

2. It is important for researchers and educators to collaborate to find creative and innovate methods for accounting for complexity when developing and testing models of long-term impact. As well, researchers and educators must define and operationalize what effects or outcomes are to be prioritized thus increasing the likelihood that educational strategies will be designed to effectively achieve those outcomes, and that the assessment of those outcomes will be aligned with the effects that the program or institution is best positioned to support

3. The field must confront the complexity of attribution by designing educational systems and partnerships that have a broader influence on the many learning contexts and experiences across a person's life and thus exert greater, more synergistic influence on long-term learning impacts.

4. Researchers must identify specific moments in the process to review and modify the process in response to the researcher's own learning, including newly identified cultural biases.

5. As the field builds improved culturally sensitivity in its research designs and practice, it will necessarily move away from prescriptive, linear approaches and, indeed, open the research process up to be more inclusive of its subjects.

These five recommendations represent important first steps towards achieving the goal of more validly and reliably measuring the long-term, cascading effects of any particular educational experience. However, further thought and experimentation on this topic are clearly needed. Additional efforts might include:

• Using these initial findings as a foundation for publications or presentations, and as a vehicle for generating further discussion;

• Encouraging other investigators to add to these initial reviews and collectively generate a more comprehensive review of existing literature, with particular attention to identifying key gaps in analysis;

• Convening of a larger, follow-up conference on this issue that includes invitations to those performing long-term research in health and wellness, human development, and other relevant social science areas; and

• The specific earmarking of funding for support of efforts to validly and reliably study long-term effects.

Acknowledgements

We gratefully acknowledge that this work was supported in part by the National Science Foundation (Pathways Grant #DRL-1515550).

References and Resources

References

Adelman, L.M., Falk, J.H., & James, S. (2000). Assessing the National Aquarium in Baltimore's impact on visitor's conservation knowledge, attitudes and behaviors. Curator, 43(1), 33–62.

Ahmed, S. & Palermo, A-G. (2010). Community engagement in research: Frameworks for education and peer review, *American Journal of Public Health*, 100(8), 1380–1387.

Anderson, D., Lucas, K., Ginns, I., & Dierking, L. (2000). Development of knowledge about electricity and magnetism during a visit to a science museum and related post-visit activities. *Science Education*, 84(5), 658–679.

Anderson, D., Storksdieck, M. & Spock, M. (2006). Long-term impacts of museum experiences. In J. Falk, L. Dierking and S. Foutz (eds.) *In Principle, In Practice*, (pp. 197–215), Lanham, MD: AltaMira Press.

Baer, J. M. (1988). Long-term effects of creativity training with middle school students. *The Journal of Early Adolescence*, 8(2), 183–193. https://doi.org/10.1177/0272431688082006

Baer, R. A., Rinaldo, J. C., & Berry, D. T. G (2003). Response distortions in self-report assessment. In R. Fernandez-Ballesteros (Ed.), *Encyclopedia of psychological assessment*. London: Sage.

Barron, B. (2006). Interest and self-sustained learning as catalysts of development: A learning ecology perspective. *Human Development*, 49(4), 153–224.

Bathgate, M.E., Schunn, C.D. & Correnti, R. (2013). Children's motivation toward science across contexts, Manner of interaction, and topic. *Science Education*, 97, 1–28.

Bell, P., Bricker, L., Reeve, S., Toomey Zimmerman, H. & Tzou, C. (2013). Discovering and Supporting Successful Learning Pathways of Youth In and Out of School: Accounting for the Development of Everyday Expertise Across Settings. *LOST Opportunities*, Springer Netherlands, pp. 119–140.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds.) (2000). *How people learn*. Washington, DC: National Research Council.

Campbell, D. T., & Stanley, J. C. (1967). *Experimental and quasi-experimental designs for research* (2nd ed.). Boston, MA: Houghton Mifflin Company.

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309–335). New York, NY: Routledge.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, 320(5880), 1217–1220.

Dumont, H., Istance, D. & Benavides, F. (Eds.) (2012). *How can the learning science inform the design of 21st century learning environments? Centre for Educational Research and Innovation*. OECD Publications. http://www.oecd.org/education/ceri/50300814.pdf Retrieved May 14, 2018.

Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: an introductory guide to latent variable growth curve modeling. *Social and personality psychology compass*, 3(6), 979–991.

Eaglemen, D. (2015). *The brain*. New York: Pantheon.

Falk, J. H. (2009). *Identity and the museum visitor experience*. Walnut Creek, CA: Left Coast Press.

Falk, J.H. (2018). *Born to Choose: Evolution, self and well-being*. London: Routledge.

Falk, J. H., & Dierking, L. D. (1997). School field trips: Assessing their long-term impact. *Curator: The Museum Journal*, 40(3), 211–218. https://doi.org/10.1111/j.2151-6952.1997.tb01304.x

Falk, J. H., & Dierking, L. D. (2010). The 95 percent solution: School is not where most Americans learn most of their science. *American Scientist*, 98(6), 486–493. https://doi.org/10.1511/2010.87.486

Falk, J.H. & Dierking, L.D. (2014). *The Museum Experience Revisited*. Walnut Creek, CA: Left Coast Press.

Falk, J.H. & Dierking, L.D. (2018). *Learning from Museums*. Lanham, MD: Rowman & Littlefield.

Falk, J. H., Dierking, L. D., & Foutz, S. (Eds.). (2007). *In principle, in practice: Museums as learning institutions*. Lanham, MD: AltaMira.

Falk, J.H., Dierking, L.D., Osborne, J., Wenger, M., Dawson, E. & Wong, B. (2015). Analyzing science education in the U.K.: Taking a system-wide approach. *Science Education*, 99(1), 145–173.

Falk, J. H., Dierking, L. D., Swanger, L. P., Staus, N., Back, M., Barriault, C., … Verheyden, P. (2016a). Correlating science center use with adult science literacy: An international, cross-institutional study. *Science Education*, 100(5), 849–876. https://doi.org/10.1002/ sce.21225

Falk, J. H., & Needham, M. D. (2013). Factors contributing to adult knowledge of science and technology. *Journal of Research in Science Teaching*, 50(4), 431–452.

Falk, J. H., Pattison, S. A., Meier, D., Bibas, D., & Livingston, K. (2018). The contribution of science-rich resources to public science interest. *Journal of Research in Science Teaching*, 55(3), 422–445. https://doi.org/10.1002/tea.21425

Falk, J.H., Scott, C., Dierking, L.D., Rennie, L.J., & Cohen Jones, M. (2004). Interactives and visitor learning. *Curator*, 47(2), 171–198.

Falk, J. H., Staus, N., Dierking, L. D., Penuel, W., Wyld, J., & Bailey, D. (2016b). Understanding youth STEM interest pathways within a single community: The Synergies project. *International Journal of Science Education*, Part B, 6(4), 369–384. https://doi.org/10.1080/21548455.2015.1093670

Falk, J. H., & Storksdieck, M. (2005). Using the contextual model of learning to understand visitor learning from a science center exhibition. *Science Education*, 89(5), 744–778. https://doi.org/10.1002/sce.20078

Farinde, A. (2017). Dose-Response Relationships. https://www.merckmanuals.com/profesional/clinical-pharmacology/pharmacodynamics/dose-response-relationships

Flagg, B. (2005). Beyond entertainment: Educational impact of films and companion materials. *Big Frame*, 22 (2), 50–56.

Fivush, R., McDermott Sales, J., Goldberg, A., Bahrick, L., & Parker, J. (2004). Weathering the storm: Children's long-term recall of Hurricane Andrew. *Memory*, 12(1), 104–118. https://doi.org/10.1080/09658210244000397

Fraser, J., Heimlich, J.E., Jacobson, J., Yocco, V., Sickler, J., Kisiel, J., Nucci, M., Ford Jones, L. & Stahl, J. (2012). Giant screen film and science learning in museums. *Museum Management and Curatorship*, 27(2), 179–195.

Fu, A. C., Kannan, A., Shavelson, R. J., Peterson, L., & Kurpius, A. (2016). Room for rigor: Designs and methods in informal science education evaluation *Visitor Studies*, 19(1), 12–38. https://doi.org/10.1080/10645578.2016.1144025

Gates, E., & Dyson, L. (2017). Implications of the changing conversation about causality for evaluators. *American Journal of Evaluation*, 38(1), 29–46. https://doi.org/10.1177/1098214016644068

Gonyea, R. M. (2005). Survey research: Emerging issues. *New Directions for Institutional Research*, 2, 73–89.

Groves, R. (1989). Survey costs and survey errors. New York: John Wiley.

Howell, R. A. (2014). Investigating the long-term impacts of climate change communications on individuals' attitudes and behavior. *Environment and Behavior*, 46(1), 70–101. https://doi.org/10.1177/0013916512452428

Isiordia, M., & Ferrer, E. (2018). Curve of factors model: a latent growth modeling approach for educational research. *Educational and Psychological Measurement*, 78(2), 203–231.

Ito, M., Baumer, S., Bittanti, M., Boyd, D., Cody, R., Herr-Stephenson, B., Horst, H.A., Lange, P.G., Mahendran, D., Martinez, K.Z., Pascoe, C., Perkel, D., Robinson, L., Sims, C. & Tripp, L. (2013). *Hanging out, messing around, and geeking out: Kids living and learning with new media*. Cambridge, MA: MIT Press.

Lavrakas, P.J. (Ed.). (2008). Encyclopedia of survey research methods. Thousand Oaks, CA: Sage.

Lemke, J. L., Lecusay, R., Cole, M., & Michalchik, V. (2015). *Documenting and assessing learning in informal and media-rich environments*. Retrieved from http://mitpress.mit.edu/sites/default/files/9780262527743%20(2).pdf

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child development*, 110–133.

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. Washington, DC: National Academies Press.

National Research Council. (2015). *Identifying and supporting productive STEM programs in out-of-school settings*. Washington, DC: The National Academies Press.

Pattison, S. A., & Shagott, T. (2015). Participant reactivity in museum research: The effect of cueing visitors at an interactive exhibit. *Visitor Studies*, 18(2), 214–232. https://doi.org/10.1080/10645578.2015.1079103

Pattison, S. A., Svarovsky, G. N., Gontan, I., Corrie, P., Benne, M., Weiss, S., … Ramos-Montañez, S. (2017). Teachers, informal STEM educators, and learning researchers collaborating to engage low-income families with engineering. *Connected Science Learning*, 4. Retrieved from http://csl.nsta.org/2017/10/head-start-engineering/

Peterman, K., Pressman, E. & Goodman, I.F. (2007). NOVA science Now science cafés evaluation. Unpublished Technical Report. NY: Goodman Research Group.

Peterson, C. (2002). Children's long-term memory for autobiographical events. *Developmental Review*, 22(3), 370–402.

Quinn Patten, M. (2016). *Developmental Evaluation*. New York: Guilford Press.

Roschelle, J. (1995). Learning in interactive environments: Prior knowledge and new experience. In: J. Falk & L. Dierking (eds.) *Public Institutions for Personal Learning* (pp. 37–51). Washington, DC: American Association of Museums.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Career choice: Planning early for careers in science. *Science*, 312(5777), 1143–1144.

Taplan, S. (2005). Methodological design issues in longitudinal studies of children and young people in out-of-home care: A literature review. Technical Report. NSW Centre for Parenting & Research. Sydney, Australia: NSW Department of Community Services.

Vaske, J. J. (2008). Survey research and analysis: *Applications in parks, recreation and human dimensions*. State College, PA: Venture.

Resources

Collins et al., 2018 Community Based Research Values and Principles (suggestions for creating an environment for meaningful community involvement in all stages of research process).

Harvard Adult Development Study http://www.adultdevelopmentstudy.org/ Retrieved September 25, 2018.

Lavery, J.V. et al. (2010). *Towards a framework for community engagement in global health research* (PubMed link: https://www.ncbi.nlm.nih.gov/pubmed/20299285) This is a public health research opinion in the Journal of Parasitology that outlines key practices in effective community engagement RE: health research, including involvement in developing research goals and paying attention to changes in populations over time. PDF link: https://www.dropbox.com/s/n22a7pi8qcp5i76/ PIIS1471492210000425.pdf?dl=0

Lincoln, Y.S. & Guba, E.G. (1985) *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.

Appendix A: Participants

Lynn Dierking is Director of Strategy & Partnerships, Institute for Learning Innovation, and Professor, Free-Choice Learning, Oregon State University. Her research on lifelong, out-of-school learning (after-school, home- and community-based contexts), with youth and families, focuses primarily on youth/families living in poverty and/or not historically engaged in free-choice learning from cultural institutions/organizations. Dr. Dierking is PI of a US-NSF project, SYNERGIES: Customizing Interventions to Sustain Youth STEM Interest and Participation Pathways, studying youths' STEM interest and participation longitudinally in an under-resourced community. She also is co-PI of a US-NSF/UK-Wellcome Trust Science Learning+ Partnership project, Partnering for 'Equitable STEM Pathways' for Youth Underrepresented in STEM. She is on Editorial Boards for Connected Science Learning, Afterschool Matters and Journal of Museum Management and Curatorship.

John Falk is Director of the Institute for Learning Innovation and Emeritus Sea Grant Professor of Free-Choice Learning at Oregon State University. He is a leading expert on free-choice learning; the learning that occurs when people have significant choice and control over the what, where and when of their learning. His current research focuses on understanding the identity/self-related reasons people utilize free-choice learning settings during their leisure time; studying the community impacts of museums, libraries, zoos and aquariums, measuring the long-term interest pathways of youth and helping cultural institutions re-think their educational positioning in the 21st century.

Gail Jones has a PhD in Science Education from NC State University. Dr. Jones currently serves as Alumni Distinguished Graduate Professor of Science Education and a Fellow at the Friday Institute for Educational Innovation teaching preservice and in-service teachers and conducting research on virtual reality, nanotechnology and family learning in out-of-school contexts. Dr. Jones' research has been recognized by the National Association for Research in Science Teaching, The NC Association of Research in Education, and the Association of Supervision and Curriculum Development. Dr. Jones has authored several books for teachers: Nanoscale Science, Extreme Science, and Case Studies in Biology and Engineering (in press). Dr. Jones' research group is currently researching new forms of technology for teaching science and strategies to enhance science capital and family habitus for science.

Jim Kisiel is Associate Professor of Science Education at California State University, Long Beach. Much of his research has examined the juxtaposition of formal and informal environments, examining the opportunities and constraints in collaborative activities ranging from field trips to more formal school-museum partnerships. He has also conducted a variety of research and evaluation studies examining different learners in informal contexts. These include clarifying adult museum-goers' understanding of science, identifying family learning behaviors, and examining science identity development.

Judith Koke is Director, Professional Learning at the Institute for Learning Innovation where she leads the Institutes efforts to research, innovate and disseminate effective methods to engage professionals in learning and capacity building, particularly with in regard to integrating research into practice. Previously as Senior Research Associate at the Institute, she has a long history of research and evaluation in the free-choice learning field. She has also worked in senior leadership roles at the Art Gallery of Ontario and The Nelson-Atkins Art Museum. With a career spent in both research and museum leadership, she understands how to assess and apply research findings into better practice.

Kate Livingston is the founder and Principal at ExposeYourMuseum LLC (Detroit, MI), a consulting firm supporting arts and cultural organizations to better understand their internal climate and culture, current and prospective audiences, and role and potential in their communities. Kate has 15+ years of experience designing, developing, and executing professional research and evaluation and 10+ years of experience designing and implementing strategic, master, and interpretive plans. Her approach prioritizes making connections, facilitating conversations, elevating communities, engaging creatively, and strong, clear communication to inspire innovation, inform strategy, and drive decision-making. Kate is committed to inclusion, anti-racism, and social justice work, and these principles are central to her work alongside museums. From 2007-2013, Kate led the department of Audience Insights at the Denver Museum of Nature & Science (DMNS).

Rabiah Mayas is the Associate Director of Science in Society at Northwestern University, where she leads the development, implementation and evaluation of K-12 STEM education programs and partnerships in Chicago and Evanston. Key areas of focus include afterschool STEM mentoring for middle-grade youth, training of STEM graduate students in community engagement, and NGSS professional development for CPS high school teachers. Prior to joining Northwestern in 2017, Rabiah was the Director of Science and Integrated Strategies at the Museum of Science and Industry, Chicago. Key program areas developed or expanded under her leadership include maker-based learning experiences, public programs of the Black Creativity initiative, and evaluation and science learning research. Rabiah completed her Ph.D. in biochemistry and molecular biology at the University of Chicago.

Ali Mroczkowski is a Researcher and Project Manager at the Museum of Science and Industry, Chicago. She earned her Ph.D. in Community Psychology from DePaul University in 2017. Her research is on the educational experiences of marginalized youth. Recently, her research has focused on the role of out-of-school time programs in supporting the educational and career development of youth.

Scott Pattison is a researcher and evaluator at TERC, formerly the Institute for Learning Innovation. Over the last 15 years, his work has focused on education, learning, and interest development in free-choice and out-of-school environments, including museums, science centers, and everyday settings. Dr. Pattison specializes in using qualitative and quantitative methods to investigate the processes and mechanisms of learning in naturalistic settings. He is committed to addressing issues of equity and inclusion in education and has partnered with organizations across the country to support learning for diverse communities.

Aaron Price is the Director of Research and Evaluation at the Museum of Science and Industry, Chicago (MSI). His team of six studies the Museum's impact on guests and the community. He earned his Ph.D. in science education (learning sciences) after working for 14 years at an astronomical citizen science organization.

Robert Tai is an Associate Professor of Science Education at the University of Virginia in the Department of Curriculum, Instruction, and Special Education. Prior to joining the faculty at the University of Virginia, Dr. Tai taught high school physics in Illinois and then Texas. He has served as both a research associate and teaching fellow in the Graduate School of Education at Harvard University. Dr. Tai is fully involved in several grant funded research projects, the supervision and mentoring of doctoral students, and the production and dissemination of science education research.