5

Evaluation Capacity Building for Informal STEM Education: Working for Success Across the Field

Marjorie Bequette, Christopher L. B. Cardiel, Sarah Cohn, Elizabeth Kunz Kollmann, Frances Lawrenz

Abstract

Informal STEM education (ISE) organizations, especially museums, have used evaluation productively but unevenly. We argue that advancing evaluation in ISE requires that evaluation capacity building (ECB) broadens to include not only professional evaluators but also other professionals such as educators, exhibit developers, activity facilitators, and institutional leaders. We identify four categories of evaluation capacity: evaluation skill and knowledge, use of evaluation, organizational systems related to conducting or integrating evaluation, and values related to evaluation. We studied a field-wide effort to build evaluation capacity across a network of organizations and found it important to address individuals' evaluation capacities as well as capacities at the organizational level. Organizational factors that support ECB included redundancy of evaluation capacities across multiple people in an organization, institutional coherence around the value of evaluation, and recognition that ECB can be led from multiple levels of an organizational hierarchy. We argue that the increasing emphasis on evaluation in the ISE field represents an exciting opportunity and that, with targeted strategies and investments, ECB holds great promise for the future of ISE and ISE evaluation. © 2019 Wiley Periodicals, Inc., and the American Evaluation Association.

The historical and current statuses of evaluation in informal science, technology, engineering, and mathematics (STEM) education (ISE) create a setting in which evaluation has been productively but unevenly used for program improvement, documentation, and knowledge sharing. In this chapter, we argue that evaluation capacity building (ECB) is key to understanding and advancing the role and possibilities—and especially the future—of evaluation in ISE, and to maximizing the broader use of evaluation to fulfill the promise of high quality educational organizations and programs. This chapter is based on our experience leading and collaborating on evaluations in ISE, situated both externally (in university and consulting settings) and internally (as employees of ISE organizations). We also draw on our experience researching an effort to build evaluation capacity in ISE. This multiyear research study examined ECB within a network of ISE organizations, providing a broader view than is typically possible. Our work is grounded in a particular realm of ISE: museums and science centers most of all, including exhibits, museum floor programs, youth development programs, professional development for science teachers, and other museum-based programs, as well as holistic, organization-wide efforts that include these programs. Our work does not include television programs, podcasts, nature centers, and many other out-of-school settings in which people engage with and learn about science, though many of the characteristics of museums described here may also be true of those settings.

Previous chapters in this issue document the tools and practices that evaluators in ISE can use to produce credible evaluation in the field. Our focus is primarily on the ways that *non-evaluators* in ISE can do and use evaluation, and the supports that can help them in doing so most productively. We base this call for hard work in the reality of ISE organizations, including their strengths and challenges, and suggest that the pay-off of improved educational experiences is worth the attention to building a climate of evaluation practice and evaluation use beyond the community of evaluators only.

What Is ECB?

Scholarship on Evaluation Capacity Building

Across the broader evaluation field, ECB has gained prominence in the last decade (King & Volkov, 2005; Preskill & Boyle, 2008). Scholars have put forth ECB as a change effort that fosters both (a) individuals' skills and knowledge to conduct evaluation and (b) organizational structures and cultures that support and value the use of evaluation (Huffman, Thomas, & Lawrenz, 2008; Preskill & Boyle, 2008). Stockdill, Baizerman, and Compton (2002) defined ECB as "the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine" (p. 14).

Although ECB is a critical topic for the future of evaluation, more in-depth understanding of how and where it can occur is needed (Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012). This call was solidified in *American Journal of Evaluation*'s Forum section (March, 2014), which provided insights into the existing state of ECB with seven articles. Leviton (2014) raised important questions about the value of ECB to organizations, evaluators, and funders. Wandersman (2014) described ten steps for "Getting to Outcomes," and Labin (2014) presented an integrated evaluation capacity building (IECB) model. Preskill (2014) pointed out that ECB should itself be evaluated, and Suarez-Balcazar and Taylor-Ritzler (2014) emphasized the need for the science and practice of ECB to be directly linked.

Over the course of our project, discussed in more depth below, we identified four kinds of evaluation capacity in the literature that we felt were important in ISE settings. They are:

- *Evaluation Skill and Knowledge*. Often the easiest to imagine, this refers to the development of individuals' knowledge about definitions of evaluation, evaluation processes or practices, and the ability to do evaluation (Preskill & Boyle, 2008; Stockdill et al., 2002; Suarez-Balcazar et al., 2010; Taylor-Powell & Boyd, 2008).
- *Use of Evaluation*. This category includes not only the growth in quantity of evaluations being conducted, but also changes in how individuals or organizations are able to use data or findings (Naccarella et al., 2007; Taylor-Powell & Boyd, 2008).
- *Systems Related to Conducting or Integrating Evaluation*. This refers to the organizational or field-wide resources and structures—for instance, a regular spot on an agenda, a job line with evaluation responsibility, or a reliable and funded source for surveys—that support evaluation practices (Arnold, 2006; King & Volkov, 2005; Preskill & Torres, 1999; Taylor-Powell & Boyd, 2008).
- *Values Related to Conducting or Integrating Evaluation*. The perceptions of what evaluation does and why it is important, how it might have influenced past work or could change future plans, why it might be needed in a particular setting: all are part of what can be changed through ECB (Huffman et al., 2008; Preskill & Boyle, 2008; Stockdill et al., 2002; Suarez-Balcazar et al., 2010; Taylor-Powell & Boyd, 2008).

These ways of supporting evaluation—the ability to do evaluation, the actual doing and using of evaluation, the systemic structures to support evaluation (or impede it, if the structures are lacking or faulty), and the sense of the importance of evaluation—are all part of a healthy, intentional ECB effort.

Evaluation and ECB in the Informal STEM Education Context

Within the informal STEM education (ISE) context, these kinds of capacities are built within a larger system that has seen dramatic changes regarding evaluation expectations, use, and culture in recent years. Within large and small ISE organizations, the importance of evaluation as a practice for internal improvement and for external reporting and accountability has increased, particularly in the last 20 years. This increase in evaluation requirements is especially striking in ISE settings because their evaluation work had historically been less formalized than in formal education and other settings.

Funding Streams Create Need for Evaluation. ISE organizations like museums are reliant on a range of funding sources, different at every site. Public funding might be significant or close to nonexistent, endowments might be large or small, and attendance might be steady year over year or vary widely. What most sites have in common, though, is a sense of both scarcity of funds overall and a constant search for additional external funds (Beetlestone, Johnson, Quin, & White, 1998).

The search for broader, varied funding may make these organizations more moldable and responsive to external needs, including funder requirements. Many ISE organizations began their involvement with evaluation to fulfill a funder requirement, and see evaluation as something they do to receive more funding. These funders include large federal agencies, described in more depth below, but also foundations and private funders who may specifically request to see documentation of results for the funding provided.

Historic Staffing Models Typically Do Not Include Evaluation as a Position or Function. Across the ISE field, few settings have made evaluation departments or even single evaluators as part of their operating staff. Large ISE organizations (including the ones where three of the authors work) host evaluation departments, but the majority of ISE organizations, especially small ones, have no evaluation staff and are dependent on external evaluators from universities or private firms (e.g., our two other authors). At sites where evaluation is not centered in a specific role, some institutions contract for evaluation while others do not. When evaluation is only done sporadically, or at the behest of particular funders, adding evaluation-dedicated staff to an organization might not make sense.

Evaluation capacity may instead be housed among the non-evaluation staff of an ISE organization. These staff might do evaluation themselves, or might contract external firms to conduct evaluation for specific projects. Either option is lower cost than adding a dedicated position but brings its own challenges. Museum staff do not have a consistent training pathway to prepare for their professional roles; licensure or certification is not required. Ongoing professional development for museum staff is not legally

mandated. An easy comparison can be made to public school teaching and medicine, where both initial licensure and ongoing certification are required. While some museum employees enter the field from graduate museum studies programs, many come from other settings and with varying educational backgrounds. Among museum education programs, some offer evaluation courses and even an evaluation track for a master's-level degree, but evaluation coursework is typically elective and limited, especially practical work in how to do evaluation (Kannan, Fu, Shavelson, & Kurpius, 2016). Further, this degree is not required for employment in a museum. Both through professional preparation and hiring expectations, there is no common expectation that museum staff have grounding in evaluation.

Evaluation Requirements and Supports Increased at a National Level. Beginning in the early 2000s, the ISE field changed norms, expectations, and assumptions around evaluation. These changes, led mostly by the National Science Foundation (NSF) but supported by others, are described more deeply by Allen and Peterman (this issue) and Grack Nelson, Goeke, Auster, Peterman, and Lussenhop (this issue). The development of shared outcomes, the increasing attention to the quality of instruments, the supports for a community of professionals involved in evaluation in ISE, and the rising demand for ISE institutions to report on their outcomes have all changed the work being done. One key resource in this area is the *Principal Investigator's Guide: Managing Evaluation in Informal STEM Education Projects*, which was developed to support non-evaluators working as Principal Investigators (PIs) on NSF grants and includes advice on how to collaborate with the project evaluators for greatest impact (Bonney, Ellenbogen, Goodyear, & Hellenga, 2011).

Audience Research and Market Research Concerns Emerge in an Era of Data. Outside of the museum field, the increasing use of data in all aspects of society, including so-called "big data," supports a heightened awareness of the possibilities of using data more broadly. Museum professionals have watched the data analysis techniques and the resulting product development and marketing strategies used by large corporations, and some have taken note. Museum leaders responsible for finance, marketing, and ticket sales have paid more attention to how data can be used to support the bottom line, as it has in for-profit companies. Several private firms, such as Reach Advisors, IMPACTS Research and Development, and Wilkening Consulting have been developing varieties of subscription-based services that provide museums with more data about their visitors.

What Makes ECB So Crucial—and So Hard—in the ISE Field?

Previous chapters in this issue have described the new supports and expectations that can improve the work done by professional evaluators in ISE,

and the potential benefits for the ISE field that could result from that work. We argue that without corresponding attention to non-evaluators, including educators, exhibit developers, leadership, and others, the field will fall short of its potential. The combination of structural factors within training pathways and organizations, and the increased expectations for evaluation practice, create a field that is ready for strong ECB efforts. We consider the following as the primary challenges to be addressed in planning this work:

1. Many people work at museums, which are only a part of the ISE field. For context, the Association of Science-Technology Centers (which includes many but not all science museums) reports 14,760 paid employees at the 150 member museums that reported statistics for 2016 (Association of Science-Technology Centers, 2017). (There are many more science museums that did not participate in this survey.) While not all of these employees would be interested in or need ECB, this number provides context for the scale of field-wide ECB efforts. In addition, staff turnover, especially in entry-level positions, can be high. Any work aiming at ECB in ISE needs to be ready to continuously advance the work of many people, with varying motivations and goals, again and again through natural turnover, across many institutions.

2. Museums are highly variable in terms of size, content focus, and type of experience, even within only museums focused on science and STEM. When considering systems that support ECB at the organizational level, there will be no one-size-fits-all system. Field-wide plans need to be accessible to museums with large and small budgets, with varied staff, with a natural history or interactive approach, and much more.

3. Many approaches for individual roles will be needed. As is the case with any large organization, evaluation needs and uses will vary depending on the stakeholder. The value of evaluation for an educator, a finance person, and a CEO will all look different. Early efforts (including our own) focused more on how educators or exhibit developers might do and use evaluation, but these could be expanded to be used with a marketer, a box office manager, or a museum leader.

4. Adaptation of robust tools designed for trained evaluators should not be done casually; greater supports will be needed for non-evaluators who look to use these resources. For instance, as noted in Chapter 3 (Grack Nelson et al., this issue), some of the online sites that catalog instruments expect high familiarity with concepts like reliability and validity—making it harder for those with less training to choose the right tool. Recall from our earlier discussion that people who work in ISE organizations typically come from diverse

educational backgrounds and may have no formal training in education or evaluation.

What Can We Learn From a Field-Wide Effort to Build Evaluation Capacity?

The example of ECB attempted in the National Informal Science Education Network (NISE Net) provides insights into ways to address some of the issues identified above. This large network of museums, formed in 2005, developed a program to support ECB across its work—this work began in 2010 and continues to date. We and others participated in varying ways in developing and researching the program of ECB.

Background on NISE Net and Its ECB Approach

NISE Net began as a "national community of researchers and informal science educators dedicated to fostering public awareness, engagement, and understanding of nanoscale science, engineering, and technology" (www.nisenet.org). It was originally funded by the National Science Foundation through two consecutive grants that extended over 10 years and amounted to over $40 million. Continuing with new funding and new content to this day, the NISE Net is one of the largest ISE initiatives ever undertaken.

The NISE Net is characterized by leadership at multiple museums and universities, working across sites and collaborating to support many museums in engaging with new concepts. Although NISE Net has a formalized leadership structure and different types of partners, it seeks to function not as a hierarchy but rather as a collaboration of like-minded professionals working to achieve NISE Net goals. Participation in NISE Net, beyond the core partners leading the Network, is voluntary. Individuals within the network represent a range of professional roles, including informal science educators, evaluators, scientists, education outreach coordinators, and more. Products of the NISE Net include educational products (like activity kits and small exhibitions), formal professional development opportunities, and, though harder to quantify, a community of organizations that see each other as resources and colleagues for solving problems. Formal and informal meetings support this community.

Beginning in 2010, the Network launched Team Based Inquiry (TBI), an intentional ECB effort developed around ISE practitioners and *not* professional evaluators. TBI involved ISE practitioners working in groups to improve their educational products and practices, and foster effective teams and organizations; it included a variety of perspectives and backgrounds that would directly contribute to an effective inquiry process. The original approach was grounded in an extensive review of literature including work on complex systems specified by Davis and Sumara (2006) and

others, and the collaborative immersion framework for building evaluation capacity that Huffman et al. (2008) developed. The TBI approach also draws from diverse theoretical perspectives, including action research (Herr & Anderson, 2005; Loucks-Horsley, Stiles, Mundry, Love, & Hewson, 2010), practitioner inquiry (Cochran-Smith & Lytle, 2009), learning-focused and participatory evaluation (King, 1998; Preskill & Torres, 1999), and ECB (Stockdill et al., 2002). More broadly, the TBI approach is aligned with the recognized need for ongoing and iterative forms of investigation to support professional learning and organizational change (Argyris & Schon, 1995; Cochran-Smith & Lytle, 2009; Davenport & Prusak, 2000; Schon, 1983). This literature supported the development of a program that was designed to build evaluation capacity in individuals, working within organizations, thereby building the organization's capacity. It was not explicitly a "train the trainer" model in which participants would then move on to train their co-workers, though that did happen in some cases.

The implementation of TBI reached a range of NISE Net community members including NISE Net groups who learned and implemented TBI as part of their ongoing work together, NISE Net partners who received training in TBI but practiced it on projects at their own institutions, and NISE Net partners who learned about TBI through webinars or brief training but may not have formally practiced the techniques. This allowed for an informative comparison between (a) the professionals who received guided training in TBI and practiced it in their own settings and (b) professionals who participated more generally in a professional network with a focus on evaluation but did not formally practice TBI.

Team Based Inquiry, as practiced from 2010 to about 2015, addressed in various explicit and implicit ways the four kinds of evaluation capacity described earlier in this chapter: *skills and knowledge* to do evaluation, ability to *use* evaluation, *systems* to support evaluation, and *value* for evaluation. These capacities were also built by ongoing conduct, discussion, and use of formal evaluation led by evaluators during the same time:

- *Evaluation Skill and Knowledge*. TBI workshops, mentoring, and cohort learning experiences focused on the development of evaluation *skills and knowledge* at a practitioner level through explicit instruction and practice. This included how to ask good questions and how to collect and analyze data in ways that are rigorous enough for those questions and fit into practitioner work. It also included instruction on how to collect direct evidence of learning (not just self-reports of learning) when possible. Chapter 2 (Fu, Kannan, & Shavelson, this issue) discusses self-reports and direct measures in more detail.
- *Use of Evaluation*. The ability to *use evaluation* was built through asking participants to do TBI in a variety of ways, and supporting them in regularly using both practitioner-led and evaluator-led evaluation. There was less formal instruction in how to use evaluation than in how to do

evaluation, though some TBI training did include opportunities for the former. Education focused on evaluation use occurred with cohort participants as they discussed ways to take action on evaluation findings and reported back on how changes unfolded within their organizations. Evaluation use was instantiated in different ways across the Network. In some settings, TBI became part of everyday practice and a great deal of practitioner evaluation was done by some and used by all. In other settings, practitioner-led evaluation was done rarely, but formal evaluator-led evaluation was used frequently. Some Network participants reported having had little practice in using evaluation.

- *Systems Related to Conducting or Integrating Evaluation*. *Systems* were discussed in formal TBI materials and in trainings—from guidance around how to plan for integrating evaluation into meetings and project timelines to considerations around who should be involved and how to manage various responsibilities—but not at the same depth as skills and knowledge. Systems were also built into the Network: time for evaluation was given on agendas, and evaluation responsibilities were assigned to both practitioners and professional evaluators.

- *Values Related to Conducting or Integrating Evaluation*. *Value* was explicitly built into TBI instruction through discussions of where and why evaluation was important. It was also built through ongoing attention to evaluation within the Network.

Learning From the TBI Effort

The TBI effort provided an opportunity to investigate the ways that intentional ECB efforts might support professionals in ISE and support organizational change at their home institutions, during a time when evaluation was emerging as a growing concern across the field. Through the Complex Adaptive Systems as a model for Network Evaluations (CASNET) study, we investigated TBI specifically and ECB generally. CASNET was a four-year project studying how ECB could be fostered within a network. The project used a multiple case study method (Stake, 2006) to examine and understand the spread of the TBI evaluation approach within the NISE Net.

The CASNET study had two foci to understand different aspects of ECB. The first focus was on NISE Net staff who used TBI within work groups—groups that were funded to create products and implement programming for the NISE Network. The second focus was on NISE Net partners who received either deep or no training in TBI. In this chapter, we share highlights from the second focus, which is of the most interest for understanding ECB in the ISE field. Further findings from this study, which used observations, detailed interviews, and document analysis, are published elsewhere (Grack Nelson et al., 2018; Lawrenz et al., 2018).

The research findings provide insights into (a) how the four kinds of evaluation capacities highlighted earlier (*skills and knowledge* to do

evaluation, ability to *use* evaluation, *systems* to support evaluation, and *value* for evaluation) were present before training in TBI was offered, (b) how the capacities developed differently for those who received training in TBI and those who did not, and (c) key institutional factors that supported ECB efforts.

Evaluation Capacity Prior to Study Participation. First, some evaluation capacity was already present among the CASNET participants prior to their participation in the study. Prior capacities varied in type and degree. Researchers found variation in *skills and knowledge* to do evaluation and ability to *use* evaluation, with more individuals reporting low confidence in their evaluation skills and low evaluation use. Few individuals reported *coherent systems* in their organizations to support evaluation. Notably, however, several participants talked of seeing an increasing *value* placed on evaluation, especially at the field-wide level. This quotation from one of the study participants was typical of what we heard:

> I think it's more museum-wide that you'd see more papers that are being written about [evaluation] …I think it's just the industry in general is taking [evaluation] on and trying it in different ways and I think that's what's exciting.

Differences in Development of Capacities Between Groups. Second, there was evidence of ECB at individual and institutional levels over the course of the CASNET project, both for those who trained in TBI and those who did not. Individuals who did not train in TBI still reported increases in their, and their institution's, *value* for evaluation and more attempts to *use* evaluation, suggesting that the field-wide conversation continued to build momentum over the course of the project. But, only those who participated in TBI training reported that they had grown their evaluation *skills* (for instance, developing appropriate instruments or analyzing data), and that their institutions were developing *systems* to support evaluation (for instance, changing job descriptions or adding new practices in a systematic way).

Institutional Supports for ECB. Third, we identified key factors within institutions that seemed to support ECB development over time, regardless of whether participants had engaged in TBI training. These factors include having more than one person develop their evaluation capacities, establishing shared values and institutional coherence around the value of evaluation, and recognizing that ECB efforts can be led from several places in an organizational hierarchy.

Redundancy of Evaluation Capacities Was Important. Within an organization, having more than one person with evaluation capacities (of several kinds) was key for the organization's overall ECB success. Davis and Sumara (2006) argue that a common language, shared responsibilities, and

similarities in skills, knowledge, and philosophies in a certain area can lead to a foundation for shared work that can positively change an organization. There were many ways that we saw people develop their evaluation capacities (including and beyond TBI training); we saw that when more than one person had developed evaluation capacities, the organization's ECB effort generally went further, yet when only one person was championing the ECB effort, it did not go far. These pairs (or more) of people were able to bring more colleagues into their efforts, build more support systems, and make more progress overall.

Shared Value for Evaluation Allowed for Greater ECB. Within the organizations we studied, institutional coherence around the *value* of evaluation influenced the extent of ECB achieved. Coherence is more than isolated individuals seeing the value of evaluation. Instead, coherence is *shared* values and a sense that individuals are working together toward shared objectives around evaluation. Perhaps the strongest display of institutional coherence and its impact on ECB came from interviews with a manager from a mid-sized museum who had not done the TBI training. Similar to many other participants' institutions, decisions about when to conduct evaluation tended to come from the Director of Education or the Executive Director. However, when evaluations were conducted, individuals at all levels were involved in the evaluation process, as described by the manager:

> [During staff meetings], we try to mix folks up in different groups, we make sure there is a diversity of departments within each of the different sections when we've done those [institutional] evaluations.

This person painted a picture of an institution that developed a strong sense of identity around evaluation. In contrast, places with low *shared value* for evaluation showed little institutional change, even if they had one or a few people who were individually passionate about evaluation.

ECB Could Be Led Successfully From Several Locations in a Hierarchy. We saw examples of successful ECB happening in some organizations where leaders established the value of evaluation and created space for individuals to develop skills and practice evaluation use. We also saw cases where individuals lower in an organizational hierarchy, but with control over an aspect of their work, were nevertheless able to develop their skills and build systems for evaluation.

Overall, these findings suggest that evaluation capacity can be built in ISE both through intentional, specific efforts like TBI and through the ongoing field-wide conversations described earlier in this chapter. But the work is hard: Recognition of the *value* of evaluation often outpaces development of other evaluation capacities, such as evaluation of skills and knowledge, the ability to use evaluation, and the systems needed to support widespread use. Value and enthusiasm for evaluation were not

enough to build institutional evaluation capacity in many of the cases that were part of our research.

Looking Ahead and Visions for the Future

Based on what we learned through carrying out and studying TBI, we suggest that ECB continues to be a potentially important tool for the ISE field to use to improve outcomes and the evaluation of outcomes. The field is ready and people are eager for intentional efforts that will support better use of evaluation—and ultimately, better learning and engagement with science at ISE institutions. As individuals who help to shape this work within our institutions and beyond, we care deeply about providing access to evaluation for many across the ISE field.

We frame the following suggestions within the challenges identified earlier. In ISE, ECB efforts need to engage many individuals, given the scale of the field. ECB efforts need to work with institutions of many types, large and small, with approaches that are appropriate for different subject matters and different internal structures. ECB efforts need to demonstrate value for different roles at informal learning organizations. And, ECB efforts need to be built for multiple audiences, including non-evaluators specifically; we cannot expect practitioners to adapt tools and resources that were developed for professional evaluators.

In addition, these visions for the future are contextualized within our experience with the TBI effort. We saw evidence for more value being placed on evaluation regardless of our interventions, likely reflecting the increasing emphasis on evaluation in the field overall. We also saw the importance of considering both the individual and the institution, as the development of individuals' skills and knowledge, their ability to use evaluation, and individual value for evaluation were not enough. Organizational factors mattered: Having multiple people developing their evaluation capacities within a single organization and having a shared vision for evaluation were key to ECB, though the impetus for change could stem from varied levels in the organization.

Given that framing, we propose six productive strategies for the future of ECB in ISE. We describe each of these strategies next.

Repeat Evaluation 101 Over and Over

Because of staff turnover, and the sheer number of staff at ISE organizations, introductory sessions continue to be important. One practitioner trained in TBI at an institution is not enough; having multiple staff members with some capacity makes a difference. Some professional evaluators may feel that the field has moved beyond these, and enthusiastic voices may clamor for more advanced opportunities, but the need for simple, introductory training remains.

Rigorously Develop Tools for Practitioner Use by Working With Practitioners

The rapidly developing tools for professional evaluators to use (see Chapters 1, 2, and 3 of this issue) are powerful, but they are intimidating for those new to evaluation. Also powerful and needed are surveys, observational protocols, and systemic supports that are easy for practitioners to implement. As an example, we suggest the current work by the Children's Museum of Pittsburgh to document learning practices during making activities (Making + Learning, n.d.; see www.makingandlearning.org). Instead of validating the observational protocol with evaluators across the field, they are asking practitioners, with evaluator support, to use it and provide feedback. It will be primarily used to prompt conversation among practitioners about the key practices to support learning through making, thus developing evaluation capacity among several people at one institution.

Create Flexible Systems

Projects that aim to build field-wide evaluation capacity need to be able to work with a variety of institutions. The Collaboration for Ongoing Visitor Experience Studies (COVES) project, described in more depth by Grack Nelson et al. (this issue) was intentionally developed to allow organizations of various types and sizes to participate. During the pilot phase, small organizations, and organizations with varied content, participated in the development of the survey instrument. As the project has grown, this is reflected in what is offered: The pricing structure for participation varies for institutions of varied budgets, and the supports for data collection can engage an institution with a dozen evaluators or with no evaluators at all and still produce systematic results. All systems built for use across the ISE field need to do the same.

Listen to the Unusual Suspects

Stories of how museums have changed practice are important to ECB efforts. However, hearing from large organizations (such as those of this chapter's co-authors) can be intimidating to those who work in places without internal evaluators. Stories from museums *without* huge (or any) evaluation departments are critical for offering diverse visions of success and encouraging change. Recent conference sessions with titles like "Evaluation: Your 'Other Duties as Assigned' and Making it Work" (Cohn, Kenner, Trainer, & Gribko, 2018,) have done so, and more discussions like these should happen (e.g., Kenner, Cohn, Deck, Pratt, & Rathbun, 2017; Steele-Inama et al., 2017).

Promote Value to Varied Audiences

An educator, a marketer, an evaluator, and a vice president of visitor experience will all do or use evaluation for different reasons and they need narratives of use that vary. This recommendation can be easily combined with the previous one—in a small institution, how do people with varied roles work together to create shared value for incorporating evaluation into their institutional system?

Include Evaluation in Accreditation With More Depth

Current accreditation programs, such as the Museum Assessment Program for Collections Stewardship, have a light touch with regard to evaluation—they ask for evidence that evaluation happens somewhere, in some way. Accreditation should include questions about who leads evaluation, how often it happens, and how they collaborate inside and outside of their institution. Further, different museum accreditation processes should align with each other, with respect to evaluation.

Concluding Thoughts

Along with these broad recommendations for the field, we are working on ECB in our own professional domains. For instance, at the Science Museum of Minnesota, the museum leadership and evaluation staff have committed to rethinking how we do and use evaluation across the entire museum. Instead of seeing evaluation as something that is optional, driven by the interest of the program developer, museum leaders (including the first author of this chapter) have worked to develop a museum-wide evaluation plan that answers broad questions of value to the entire institution. This project involves ECB at all levels, as we work to make evaluation a part of everyone's practice. Similarly, the Oregon Museum of Science and Industry has taken steps to establish ongoing internally-funded evaluation systems to inform and improve both programs (e.g., summer camps, traveling programs) and the general visitor experience. While these systems are maintained and facilitated by evaluation staff, they involve a wide range of stakeholders representing multiple areas and levels of the museum, including senior leadership. Data collection instruments are iteratively reviewed and refined to be responsive to changing stakeholder needs. Through shared data collection platforms, museum leadership and other staff have collaborated with the evaluation team to answer questions of immediate relevance and utility while simultaneously building their comfort and familiarity—their *capacity*—with regard to evaluative thinking and planning. Both institutions have used TBI as part of the overall toolkit during this work.

This rethinking of evaluation practices is hard to carry out, but it is important to do if our organizations and our field are to reach their full potential. The practices and tools described elsewhere in this issue are important

approaches to improving ISE evaluation practice and helping to identify and support better STEM learning. But these advances will not reach their full potential without attention to the practitioners and how they engage with evaluation, which we frame as ECB. Staff at many informal learning institutions recognize the value of evaluation in supporting their work, but they need help in achieving that vision. Evaluators and leaders in the field need to work broadly to enable powerful practice, together.

References

Argyris, C., & Schon, D. A. (1995). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison-Wesley Publishing Co.

Arnold, M. (2006). Developing evaluation capacity in Extension 4-H field faculty: A framework for success. *American Journal of Evaluation*, 27(2), 257–269.

Association of Science-Technology Centers. (2017). *2016 ASTC science center statistics*. Washington, DC: Author. Retrieved from http://www.astc.org/wp-content/uploads/2017/09/ASTC_SCStats-2016.pdf

Beetlestone, J. G., Johnson, C. H., Quin, M., & White, H. (1998). The science center movement: Contexts, practice, next challenges. *Public Understanding of Science*, 7(1), 5–26.

Bonney, R., Ellenbogen, K., Goodyear, L., & Hellenga, R. (2011). *Principal Investigator's Guide: Managing evaluation in informal STEM education projects*. Washington, DC: Center for Advancement of Informal Science Education & Association of Science-Technology Centers. Retrieved from http://informalscience.org/sites/default/files/caisevsapi_guide.pdf

Cochran-Smith, M., & Lytle, S. L. (2009). *Inquiry as stance: Practitioner research for the next generation*. New York: Teachers College Press.

Cohn, S., Kenner, M., Trainer, L., & Gribko, T. (2018, May). *Evaluation: Your 'other duties as assigned' and making it work*. Session presented at the American Alliance of Museums annual meeting & MuseumExpo, Phoenix, AZ.

Davenport, T. H., & Prusak, L. (2000). *Working knowledge: How organizations manage what they know*. Boston, MA: Harvard Business School Press.

Davis, B., & Sumara, D. J. (2006). *Complexity and education: Inquiries into learning, teaching, and research*. New York: Routledge.

Grack Nelson, A., King, A. J., Lawrenz, F., Reich, C., Bequette, M., … Francisco, M. (2018). Using a complex adaptive systems perspective to illuminate the concept of evaluation capacity building in a network. *American Journal of Evaluation*. https://doi.org/10.1177/1098214018773877

Herr, K., & Anderson, G. L. (2005). *The action research dissertation: A guide for students and faculty*. Thousand Oaks, CA: SAGE.

Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation*, 29(3), 358–368.

Kannan, A., Fu, A., Shavelson, R., & Kurpius, A. (2016). *Evaluation training in informal education*. Menlo Park, CA: SK Partners Research Report.

Kenner, M., Cohn, S., Deck, L., Pratt, M., & Rathbun, M. (2017). *Small, but mighty: Successes and struggles in building evaluation capacity in small museums*. Session presented at the Association of Science-Technology Centers annual conference, San Jose, CA.

King, J. A. (1998). Making sense of participatory evaluation practice. In E. Whitmore (Ed.), *Understanding and Practicing Participatory Evaluation. New Directions for Evaluation*, 80, 57–67.

King, J. A., & Volkov, B. (2005). A framework for building evaluation capacity based on the experiences of three organizations. *CURA Reporter*, *35*(3), 10–16.

Labin, S. N. (2014). Developing common measures in evaluation capacity building: An iterative science and practice process. *American Journal of Evaluation*, *35*(1), 107–115.

Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, *33*(3), 307–338.

Lawrenz, F., Kollmann, E. K., King, J. A., Bequette, M., Pattison, S., Grack Nelson, A., … Francisco, M. (2018). Promoting evaluation capacity building in a complex adaptive system. *Evaluation and Program Planning*, *69*, 53–60.

Leviton, L. C. (2014). Some underexamined aspects of evaluation capacity building. *American Journal of Evaluation*, *35*(1), 90–94.

Loucks-Horsley, S., Stiles, K. E., Mundry, S., Love, N., & Hewson, P. W. (2010). *Designing professional development for teachers of science and mathematics* (3rd ed.). Thousand Oaks, CA: Corwin Press.

Making + Learning. (n.d.). *Making + learning*. Retrieved from www.makingandlearning.org

Naccarella, L., Pirkis, J., Kohn, F., Morley, B., Burgess, P., & Blashki, G. (2007). Building evaluation capacity: Definitional and practical implications from an Australian case study. *Evaluation and Program Planning*, *30*(3), 231–236.

Preskill, H. (2014). Now for the hard stuff: Next steps in ECB research and practice. *American Journal of Evaluation*, *35*(1), 116–119.

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, *29*(4), 443–459.

Preskill, H., & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: SAGE.

Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Stake, R. E. (2006). *Multiple case study analysis*. New York: Guilford Press.

Steele-Inama, M., Cohn, S., Israel, E., Kollmann, E., Livingston, K., & Levinsky-Raskin, S. (2017, May). *Getting started on evaluation methods*. Session presented at the American Alliance of Museums annual meeting & MuseumExpo, St. Louis, MO.

Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. In M. Baizerman, D. W. Compton, & S. H. Stockdill (Eds.), *The Art, Craft, and Science of Evaluation Capacity Building. New Directions for Evaluation*, *93*, 7–25.

Suarez-Balcazar, Y., & Taylor-Ritzler, T. (2014). Moving from science to practice in evaluation capacity building. *American Journal of Evaluation*, *35*(1), 95–99.

Suarez-Balcazar, Y., Taylor-Ritzler, T., Garcia-Iriarte, E., Keys, C. B., Kinney, L., Ruch-Ross, H., & Curtin, G. (2010). Evaluation capacity building: A culturally- and contextually grounded interactive framework and exemplar. In F. Balcazar, Y. Suarez-Balcazar, T. Taylor-Ritzler, & C. B. Keys (Eds.), *Race, culture and disability: Rehabilitation science and practice* (pp. 307–324). Sudbury, MA: Jones and Bartlett.

Taylor-Powell, E., & Boyd, H. (2008). Evaluation capacity building in complex organizations. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program Evaluation in a Complex Organizational System: Lessons from Cooperative Extension. New Directions for Evaluation*, *120*, 55–69.

Wandersman, A. (2014). Getting to outcomes: An evaluation capacity building example of rationale, science and practice. *American Journal of Evaluation*, *35*(1), 100–106.

MARJORIE BEQUETTE, PhD, is Director of Evaluation and Research in Learning at the Science Museum of Minnesota, and her research interests include professional development in museums, evaluation capacity building, and organizational change toward equity and inclusivity in museums.

CHRISTOPHER L. B. CARDIEL, MA, is the Research, Evaluation, and Impacts Manager at the Oregon Museum of Science and Industry in Portland, OR, and a doctoral candidate in Educational Leadership at Portland State University.

SARAH COHN, MA, is a consultant with non-profit organizations focusing on evaluation, organizational leadership development, and strategic planning, where her practice focuses primarily on evaluation capacity building, reflective practice, and strategy development.

ELIZABETH KUNZ KOLLMANN, ALM, is Manager, Research and Evaluation at the Museum of Science, Boston and leader of the NISE Network evaluation team, where her research interests include public engagement with science and science communication as well as evaluation capacity building.

FRANCES LAWRENZ, PhD, is a Wallace Professor of Teaching and Learning and Associate Vice President for Research at the University of Minnesota.