

A Model for the Development of Evaluation Literacy in Informal Science Educators



Joy Kubarek^a and Brian Johnson^b

a. Inform Evaluation & Research, Massachusetts, USA and Museum Studies, Department of Education, Tufts University, Massachusetts, USA; b. Inform Evaluation & Research, Massachusetts, USA

Abstract

Informal science educators are expected to understand scientific content and processes, implement effective instructional strategies, and accommodate a diverse range of audiences. Missing from this repertoire of knowledge and skills is evaluation, even though this is the primary mechanism for making evidence-based decisions to inform program design and instruction. From demonstrating impact to informing practice, evaluation is just as essential for informal science educators as assessment is to classroom teachers. This article reflects on the implementation of three separate evaluation capacity building (ECB) efforts at informal science learning organizations in order to inform a new model for developing informal science educators' evaluation literacy. The three contexts included conducting ECB with an entire department, across an organization, and working with representatives of six organizations collaborating in a network. The design of each effort was informed by Preskill and Boyle's Multidisciplinary Model of Evaluation Capacity Building (2008) and Taylor-Powell and Boyd's (2008) framework for ECB. A retrospective synthesis of formative and summative data coupled with facilitator reflections identified salient features across the three contexts to inform a model for the future. This new model represents an ECB *process* rather than a one-and-done intervention to fully develop educators' evaluation literacy. It also acknowledges the cyclical nature of learning, applying, and refining ones' understanding and skills. The model serves as a tool for facilitators of ECB to apply in informal science contexts.

Keywords: evaluation capacity building, informal science, educators

Introduction

The pedagogical expectations of informal science educators include competency in effective science instruction, knowledge of scientific practices and processes, un-

derstanding of discipline-specific content, and an ability to accommodate diverse audiences. Absent or diminished from most discourse on the competencies informal science educators should develop is evaluation (Busch, 2020; Carleton-Hug & Hug, 2010). Yet, evaluation is becoming increasingly essential for informal science institutions. In its report, *Convening on Building Capacity for Evaluation in Informal Science, Technology, Engineering, and Math (STEM) Education*, the Center for the Advancement of Informal Science Education (CAISE) detailed the results of a convening focused on improving the quality of evaluation in informal science education. The report stated “evaluation produces evidence that is critical to improving our work, driving innovation, and making the case for the outcomes and impacts of informal STEM education (ISE)” (Ellenbogen, 2014, p. 1). Likewise, the National Science Foundation (NSF)-supported Informal STEM Learning Professional Competency Framework identified “evidence-based practice” as job-specific expertise practitioners should develop over time (Hunter et al., 2018). It has become apparent that evaluation is not just a nicety, but a necessity in the practice of informal science educators.

Evaluation provides informal science educators with data to inform their practice, demonstrate impact, and affirm their value in the learning landscape. Furthermore, more funders and community members are moving past outputs and asking for evidence of outcomes. As informal science institutions face questions of how they are addressing issues of diversity, equity, and inclusion, evaluation can be the tool to help organizations better understand how they are situated in their communities and what audiences expect and value. Despite the importance of evaluation, it is often on the periphery of practice for informal science educators. Lack of resources, lack of time, and lack of expertise are cited as barriers to engaging in evaluation (Carleton-Hug & Hug, 2010; Fu et al., 2016; Khalil et al., 2017; King et al., 2015; Luebke & Grajal, 2011).

Some informal science educators have recognized the need to develop their evaluation literacy. These individuals are commonly left to seek out their own professional development, which may be piecemeal and disconnected from their day-to-day practice. Examples of teams, departments, or organizations engaging in evaluation professional development are less common. However, these efforts are typically more effective in cultivating a community of sustained evaluation practice. Leaders of these teams are expected to find such professional development on their own and with limited resources and guidance. It has become apparent that for informal science educators to successfully develop their evaluation literacy, more attention needs to be paid to promising evaluation capacity building (ECB) models specific to this professional context. An ECB model for informal science educators should draw upon effective practices in ECB while acknowledging the complex nature of informal learning—one with diverse, sometimes ephemeral audiences and experiences in a myriad of learning environments.

This article explores the application of two ECB models in informal science learning settings. The ECB interventions took place in three separate contexts all within informal science learning settings: an education department, an entire organization, and across six organizations working together as a network. Synthesis of participant data, facilitator observations, and facilitator self-reflection are used to identify the salient features of a promising ECB model for informal science educators.

Background

To consider a mechanism for developing informal science educators' evaluation literacy, it is first important to understand the professional practice of informal science educators as well as effective ECB models. For the purpose of this article, the term informal science education is used to encompass both informal science and environmental learning contexts. The two fields draw from common practices and oftentimes have an intersection of content. This is not to diminish the unique nature of teaching and learning about science versus the environment, but rather to unite them in a common end goal of developing evaluation literacy.

The Practice of Informal Science Educators

Since 1970 when the Association for Science and Technology Centers (ASTC) was first established, the field of informal science has grown exponentially; the number of organizations who are members of ASTC has increased from only 16 at its genesis to more than 600 today (Ucko, 2010). Informal science education includes learning environments such as science centers, zoos, aquariums, science museums, botanic gardens, after school clubs, online learning, and other forms of multimedia (National Research Council, 2009). As the breadth of informal science learning environments has expanded, so too have the number and diversity of informal science educators. From interpretation to exhibit design to facilitating formal school programs, informal science educators take on a wide range of roles and responsibilities. However, the majority of informal science educators lack formal training in their profession and often use a learn-as-you-go approach to honing their craft (Castle, 2006). While the literature on the nature of informal science learning has grown, there is still very little research on the professional practice and subsequent development of competencies for informal science educators (Allen & Crowley, 2014; Ash, Lombana, & Alcala, 2012; Tran, 2007).

Tran and King's (2007) work laid the foundation to formalize the practice of informal science educators. Tran and King proposed three different domains of knowledge for museum educators: museum content knowledge, museum pedagogical knowledge, and museum contextual knowledge. The Lawrence Hall of Science subsequently developed and continues to implement the Reflecting on Practice (RoP) professional learning series for informal science educators (Tran & Halverson, 2020). RoP is one of the few professional development offerings specific to informal science educators. The series includes four modules: (1) learning, reflections, and science; (2) how people learn; (3) learning conversations; and (4) objects and design. However, RoP still lacks an explicit focus on developing evaluation (or assessment) competencies. Likewise, Patrick (2017) edited a handbook on the preparation of informal science educators titled *Preparing Informal Science Educators: Perspectives from Science Communication and Education*. The handbook includes chapters on defining informal science education, professional development, designing programs, bridging formal and informal education, and science communication. Once again absent is any reference to informal science educators supporting or conducting evaluation.

More recently, Busch (2020) proposed a contemporary framework for infor-

mal science educators pedagogical content knowledge (iPCK) that includes evaluation as a core competency. Busch drew from the theoretical underpinnings of pedagogical content knowledge of formal educators. This informed a draft framework, which was then vetted and further refined through interviews with informal science education professionals and then a delphi study with iterative surveying of a broader sample of professionals. The final iPCK framework consists of knowledge of informal science education goals, knowledge of informal science education programs, knowledge of contexts of audiences, knowledge of engagement strategies, and knowledge of evaluation. However, in the delphi study, respondents ranked evaluation as least important. Open-ended comments suggested professionals did not prioritize evaluation skill building and subsequent practice because it was perceived as difficult and beyond their day-to-day duties. In another study by Busch, Tate-Stevenson, Green and Chesnut (2019), social network analysis explored how environmental educators across the state of North Carolina shared goals, practices, and other forms of support related to five areas: (1) Administration; (2) Evaluation; (3) Best Practices; (4) Diversity, Equity, and Inclusion; and (5) Complex Science Topics. Once again, evaluation was the least discussed topic across educators.

Our understanding of the practice of informal science educators continues to evolve as more attention is paid toward this profession. However, this understanding will remain incomplete if competency in evaluation continues to be neglected. In formal education, assessment (the equivalent of evaluation) is part and parcel to the practice of classroom educators. Educators are expected to understand how to assess their students' learning and use this information to modify their instruction. Teacher training programs include assessment coursework and school districts support professional development focused on assessment. While the work of informal science educators has its divergences from formal education, understanding what is effective instruction should be universal across both sectors.

Evaluation in Informal Science Institutions

Evaluation in informal science institutions has certainly progressed in recent years, but there is still substantial room to grow. Carleton-Hug and Hug (2010) conducted a review of literature to identify challenges and opportunities for evaluation within environmental education. First, they found a lack of clear program objectives. Clear program objectives are necessary to identify intended outcomes and to develop an evaluation plan to measure those outcomes. Second, they found that the majority of evaluations were summative in nature with few front end or formative evaluations reported. Methods were also quite traditional and limited in nature, heavily relying on surveys and pre-post tests. Lastly, they found the evaluations ignored the "unique contextual socio-political factors relevant to each program" (p.162). The shortcomings Carleton-Hug and Hug uncovered in this review may be indicative of a larger problem--a lack of evaluation literacy within informal learning professions.

Despite the challenges highlighted by Carleton-Hug and Hug, there has been some progress in bringing evaluation to the forefront of conversation within informal learning professions. For example, the Journal of Museum Education published an issue dedicated to empowering educators to evaluate (Kubarek & Trainer, 2015). The special issue on evaluation included topics such as ethics, data

collection methods, and capacity building approaches (Kubarek, 2015).

Funders, practitioners, and evaluators alike have recognized the hurdles that must be overcome to further develop the evaluation capacity of the informal learning field. The previously mentioned convening of informal science professionals about building capacity for evaluation identified three key areas to address: shared measurement, access to resources, and professional development (Ellenbogen, 2014). Much of the progress made since the convening has focused on resources. For example, CAISE established the website informalscience.org, a repository of evaluation and research resources for informal science learning. The Institute for Museum and Library Services (IMLS) provides access to evaluation resources on its website and specifically calls out evaluation as an area of focus for professional development in the Museums Empowered grant program. The University of Michigan hosts the *My Environmental Education Evaluation Resource Assistant (MEERA)* website to provide evaluation resources and “how-to’s” for environmental education professionals (Zint, 2020).

Shared measurement, however, has gotten little traction with few successful examples. One notable example focused on visitor studies: the Collaboration for Ongoing Visitor Experience Studies (COVES) hosted by ASTC. The National Oceanographic and Atmospheric Administration (NOAA) has also established standards for meaningful watershed education experiences (MWEES), which includes outcomes in common, though these are primarily focused on a specific type of watershed programming.

Professional development has also received little attention. However, some associations have made efforts to provide support for practitioners. For example, the North American Association for Environmental Education (NAAEE) includes a research and evaluation section within its eePro e-learning series and has begun to host evaluation clinics at its annual conference. The Association for Zoos and Aquariums (AZA) recently established a Social Science Research and Evaluation advisory group to address the growing need for connecting practitioners to research and evaluation.

While the access to resources and growing visibility of evaluation within the field of informal science education is promising, this does not replace the knowledge and skills development required to implement evaluation in these settings. Establishing ECB models tailored to the unique nature of informal science learning can help bridge that gap.

Evaluation Capacity Building

ECB efforts span a number of disciplines, from social services to health care to education. Labin et al. (2012) define ECB as “an intentional process to increase individual motivation, knowledge, and skills, and to enhance a group or organization’s ability to conduct or use evaluation” (p.2). The key term within this definition is intentional. While practitioners may be involved in an evaluation process as a stakeholder, this does not replace explicit knowledge and skill building efforts to equip them to conduct or support evaluation. Involvement in an evaluation study may help raise awareness and buy in, but on its own is not sufficient to develop evaluation literacy.

ECB has gained attention over the past 20 years as collaborative approaches

to evaluation have evolved. ECB works with individuals and organizations “to enhance stakeholders’ understanding of evaluation concepts and practices” and to “help [organization] members learn about and engage in evaluation practices” (Preskill & Boyle, 2008, p. 443). While there are a number of ECB models, the work highlighted in this article applied core elements of two models: Preskill and Boyle’s Multidisciplinary Model for Evaluation Capacity Building (2008) and Taylor-Powell and Boyd’s (2008) framework for ECB from Cooperative Extension Systems.

The ECB model proposed by Preskill and Boyle, in particular, has been highly discussed and applied in the professional literature. Preskill and Boyle’s model is extensive with concentric spheres of interrelated goals, factors, feedback, and implementation strategies to consider (Figure 1). Goals include the primary aim of the ECB effort and are strongly influenced by factors such as motivations, assumptions, and expectations. This subsequently influences which learning strategies to implement and how transfer of learning is adopted into sustainable practice. For example, an ECB effort can target a range of outcomes depending on the motivations of those involved as well as any previous experience and engagement with evaluation. Preskill and Boyle also identify 10 teaching and learning strategies for ECB but caution that these strategies must be carefully selected to fit the needs, motivations, and current capacity of the individual and/or organization. Preskill and Boyle suggest assessing the evaluation competence of potential ECB participants *before* designing the effort so as best to match these strategies with the intended outcomes and operating context. Not all of the learning strategies need to be implemented at the same time in a single ECB effort; the readiness and motivation of the organization, team, or individual embarking on the ECB journey should drive the specific ECB plan. This is an important point that will be revisited later in the discussion of how these ECB efforts work in practice with informal science learning organizations.

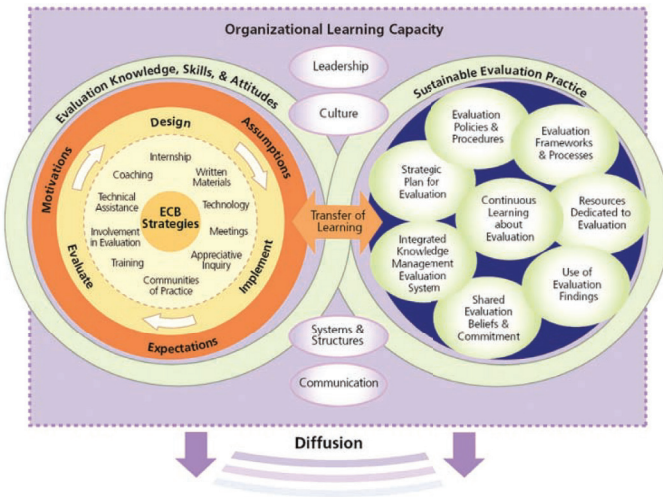


Figure 1. Multidisciplinary Model of Evaluation Capacity Building (Preskill & Boyle, 2008, p. 445)

Taylor-Powell and Boyd (2008) shared a similar but pared down model for ECB with three main components: professional development, resources and support, and organizational environment. Within each of these components there is significant overlap with the teaching and learning strategies identified in Preskill and Boyle's (2008) model. For example, professional development consists of training, technical assistance, collaborative evaluation projects, mentoring and coaching, and communities of practice. All of these activities are present in Preskill and Boyle's (2008) model, as well. Resources and support consist of evaluation and capacity building expertise, materials, evaluation champions, organizational assets, financing, technology, and time. Lastly, the organizational environment includes leadership, demand, incentives, structures, and policies and procedures. Much of Taylor-Powell and Boyd's model was informed by the authors' practical experience carrying out ECB across Cooperative Extension offices and programs.

Taylor-Powell and Boyd's (2008) model emphasized ECB as a process; it is not a one-and-done workshop with a subset of individuals. It is an organizational journey and moves beyond simply training staff. The authors stated it is important to "understand ECB as organizational development, not just professional development" (p.66). They also cautioned that not every staff member needs to become an expert evaluator. Rather, it is about cultivating evaluative thinking and an evaluation culture within the organization, one that is ready and willing to accept evaluation as integral to the organization's success. While both Preskill and Boyle (2008) and Taylor-Powell and Boyd's (2008) models include training, technical assistance, and resources, Taylor-Powell and Boyd recommend "using every opportunity, every teachable moment, and every serendipitous occasion to build evaluation capacity building" (p.67). While there is a technical learning component of ECB, it is just as important to use informal opportunities to identify and apply evaluative thinking in day-to-day practice.

Lastly, Taylor-Powell and Boyd raise an important point about who conducts the ECB process. They contend that program evaluators and ECB practitioners may not necessarily be one and the same. Rather, they characterize evaluation capacity builders as "evaluation educators." This is not to say that a professional program evaluator cannot also facilitate ECB, but it does demand a different skill set, one more akin to educators. Facilitating ECB involves teaching for knowledge and skill development, assessing for comprehension and mastery, and continued coaching of individuals. This requires not just technical prowess but also interpersonal skills to connect with others and engage in active listening.

ECB Implementation in Informal Science Learning Contexts

The following sections examine how the authors implemented core elements from both Preskill and Boyle's and Taylor-Powell and Boyd's models in three distinct informal science learning contexts: an education department (department-wide), across an entire organization (organization-wide), and with members of a regional network of organizations (network-wide). The same facilitators implemented each ECB process, ensuring consistency in facilitation as well as an opportunity to analyze successes and failures across the different contexts.

Context 1: Department-Wide

The head of an education department of a large urban zoo in the southern United States sought out the authors to facilitate an ECB process as an integral part of the organization's strategic priorities. A strategic planning process at the zoo identified the goal to "create a culture of evaluation-based decision-making" within the education department. Within this goal, specific strategies included conducting evaluation of hallmark programs, establishing a framework of outcomes in common for programs, and developing staff's ability to implement evaluation. The ECB effort was bolstered by aligning with a number of variables in Preskill and Boyle's and Taylor-Powell and Boyd's ECB models. First, as an identified strategic goal, the need, motivation, and buy-in for ECB was solidified. Leadership supported the endeavor and other key stakeholders at the zoo were aware and supportive of it as well. Second, the effort was led by an "evaluation champion," as highlighted by Taylor-Powell and Boyd. The head of the education department was enthusiastic about the opportunity and catalyzed the process. Other key elements in Taylor-Powell and Boyd's model were also present in this context, including demand, time, and financial support.

The authors first met with the head of the department to discuss the departments' needs and the end goal for this work. This helped inform the design of the ECB initiative including training topics, sequence, and format for projects. The facilitators also worked with the department's leadership team to develop a vision statement for the evaluation culture they wanted to work toward. This would serve as an anchor of the ECB process.

The ECB process spanned two and a half years and engaged all 12 staff in the department. The first activity was the development of a framework of conservation education outcomes. The framework then became the basis for program-specific logic models and evaluation plans. Once the vision statement, framework, and logic models were established, the facilitators began a series of trainings covering core evaluation competencies with a focus on case studies from similar informal learning settings. After the first cycle of training, staff then engaged in small-scale evaluation projects with coaching from the ECB facilitators. The training, participation in evaluation projects, and coaching are all learning strategies present in both Preskill and Boyle and Taylor-Powell and Boyd's models. Lastly, the facilitators created an online evaluation resource hub for staff to access training documents, resources, templates, and evaluation project updates. This accounted for the technology and written materials recommendations of both ECB models. Finally, while the training and coaching was taking place, the ECB facilitators were leading multiple program and exhibit evaluations at the zoo, offering opportunities for additional staff learning. A summary of the implementation of the department-wide ECB effort is detailed in Table 1.

Table 1. Department-Wide ECB Activities

| <i>Activity</i> |
|---|
| Conservation education outcomes framework development |
| Articulation of evaluation vision statement |

| |
|--|
| Creation of evaluation resource hub |
| Nine evaluation trainings with staff |
| Regular department leadership meetings |
| Five evaluation projects with coaching |
| Three program evaluations and 3 exhibit evaluations |
| On-call technical assistance including in-person office hours and virtual sessions |

Context 2: Organization-Wide

The second ECB effort focused on laying the foundation for evaluation across an entire organization. The organization's core work was animal rescue and rehabilitation with a strong education presence from its base on the outskirts of a large west coast city. The request for ECB was concurrent with broader organizational development efforts. The champion for the ECB effort was the head of the education department, though the intent was to work across all departments including veterinary science, rescue and response, human resources, and development.

The ECB plan was designed to span two years and consisted of needs assessments, training, coaching, evaluation projects, systems development, and a resource hub. However, implementation needed to be adaptive to a changing organizational climate and subsequently carried over to a third year. The ECB process began with creating a vision statement for the organization's evaluation culture and identifying organization-wide outcomes. Rather than a framework, the facilitators created an outcomes map detailing the organization's mission, intended impacts, subsequent outcomes, and strategies to achieve those outcomes. The organization decided to take a top-down rather than a bottom-up approach for the vision and outcomes mapping process and focused on working with the senior leadership team.

In the second year, the facilitators launched three small-scale evaluation projects working with a subset of staff. This helped fulfill the learning strategy of engaging in an evaluation process and coaching of staff. Formal training was delayed until year three because of internal changes and other current events impacting staff capacity. There were also plans to create an online evaluation resource hub for staff, just as there was in the department-wide effort. The full extent of ECB activities are detailed in Table 2.

Table 2. Organization-Wide ECB Activities¹

| |
|---|
| <i>Activity</i> |
| Organization-wide outcomes map created |
| Articulation of evaluation vision statement |

¹ The organization-wide evaluation capacity building effort was still underway at the time of publication. Year 3 activities had not yet been completed.

| |
|--|
| Regular department leadership meetings |
| Creation of evaluation resource hub (in progress) |
| On-call technical assistance |
| Three evaluation projects with coaching |
| Four evaluation trainings with staff (in progress) |

Context 3: Network-Wide

The last context involved representatives of six organizations who frequently collaborated on regional environmental education efforts in a rural area of the Pacific Northwest. One of the organizations spearheaded the effort through a successful grant award from the Institute for Museum and Library Services. The effort focused on increasing the capacity of individual organizations to conduct evaluation on their own, as well as collectively across the community. In total, the effort spanned eighteen months with the majority of the time dedicated toward individual evaluation projects and coaching. At the beginning of the ECB process, representatives from each organization participated in a three-day training to develop evaluation knowledge and skills, and identify common outcomes. Because this ECB effort involved six organizations, the group identified outcomes in common rather than develop a single framework as in the other contexts.

One distinguishing feature of the network-wide context was the use of “critical friends.” Each of the participating individuals worked with a “critical friend,” a fellow participant to engage with for peer support. The expectation was for critical friends to ask questions about their evaluation projects, help find resources when needed, provide feedback about data collection methods and instruments, and offer other general support throughout the evaluation process. Each pair of critical friends decided on the frequency and nature of these interactions. As a result, some critical friend pairs were more engaged than others. In addition to the critical friends, participants engaged in monthly learning circle meetings with the facilitators to share progress on evaluation projects, learn more about a specific evaluation topic, or seek technical support from one another and the facilitators. The learning circles fulfilled the communities of practice components of both Pre-skill and Boyle and Taylor-Powell and Boyd’s models. The full range of ECB activities is shown in Table 3.

Table 3. Network-Wide ECB Activities

| |
|---|
| <i>Activity</i> |
| Articulation of evaluation vision statement |
| Identification of common outcomes |
| Evaluation trainings with staff |
| Six organization-specific evaluation projects |

| |
|--------------------------------------|
| Individual coaching sessions |
| Critical friend peer-to-peer support |
| Monthly learning circle meetings |
| On-call technical assistance |

Comparison of Implementation

Each of the ECB efforts was catalyzed by a member of the education department, though some efforts included work beyond education programs. A retrospective review of the three ECB contexts compared with Preskill and Boyle's and Taylor-Powell and Boyd's models showed a number of activities in common as well as some divergences (Table 4 and Table 5). In these tables, the shaded boxes represent ECB activities that were present in each context.

Key features in common across the ECB contexts included training, involvement in the evaluation process, mentoring/coaching, technical assistance, written materials, financial support, and facilitators with both evaluation and ECB expertise. In each context, the involvement in an evaluation process was specific to an informal learning program and audience at that organization. Training also focused on sharing case studies from other informal learning settings. However, implementation varied for a number of reasons. First, the ECB efforts were tailored to the needs and motivations of the participating members. This subsequently influenced training topics and duration as well as how evaluation projects were conducted with staff. Second, organizational readiness and internal capacity to engage in the ECB process influenced timing and continuity of activities. The department-wide context had the most significant investment of time and staff engagement compared with the other two contexts.

Table 4. Comparison of ECB Implementation - Preskill & Boyle

| <i>Activities</i> | <i>Department-Wide</i> | <i>Organization-Wide</i> | <i>Network-Wide</i> |
|--------------------------------------|------------------------|--------------------------|---------------------|
| Internship | | | |
| Written Materials | | | |
| Technology | | | |
| Meetings | | | |
| Appreciative Inquiry | | | |
| Communities of Practice | | | |
| Training | | | |
| Involvement in an evaluation process | | | |

A Model for the Development of Evaluation Literacy

| | | | |
|-----------------------|--|--|--|
| Technical assistance | | | |
| Coaching or mentoring | | | |

Note: Shaded cells represent ECB activities that were present in this intervention.

Table 5. Comparison of ECB Implementation - Taylor-Powell & Boyd

| <i>Activities</i> | <i>Depart- ment-Wide</i> | <i>Organiza- tion-Wide</i> | <i>Net- work-Wide</i> |
|-----------------------------------|------------------------------|--------------------------------|---------------------------|
| Training | | | |
| Technical Assistance | | | |
| Collaborative Evaluation projects | | | |
| Mentoring and coaching | | | |
| Communities of practice | | | |
| Evaluation and ECB expertise | | | |
| Evaluation materials | | | |
| Evaluation champions | | | |
| Organizational assets | | | |
| Financing | | | |
| Technology | | | |
| Time | | | |
| Leadership | | | |
| Demand | | | |
| Incentives | | | |
| Structures | | | |
| Policies and procedures | | | |

Note: Shaded cells represent ECB activities that were present in this intervention.

Data Collection and Facilitator Reflection

Recognizing the opportunity to analyze successes and challenges within each context, the facilitators carried out formative and summative evaluation of each ECB effort. However, it was not the original intent to compare the three efforts, and so

the evaluation carried out for each varied. Each context occurred independently and at different points in time. The following table details the data collected in each context.

Table 6. ECB Data Collection

| <i>Context</i> | <i>Data Collected</i> | <i>Sample Size</i> |
|-----------------------------------|--|--------------------|
| 1. Department-wide | Needs Assessment Survey Formative Questionnaire Group Interviews Leadership Reflections | n=12 |
| 2. Organization-wide ² | Needs Assessment Interviews | n=8 |
| 3. Network-wide | Needs Assessment Survey Formative and Summative Surveys Individual Interviews | n=6 |

Needs Assessment

All three contexts collected needs assessment data at the onset of the ECB process. For two of the contexts (department-wide and network-wide), this took the form of a survey including questions based on the stages of change in Prochaska's Trans-theoretical Model of Behavior Change (Prochaska, Johnson & Lee, 2009). For the organization-wide effort, the needs assessment was carried out via in-person participatory methods and interviews with key stakeholders.

The questions based on Prochaska's model of behavior change were adapted for an evaluation focus, with participants self-rating their current state of practice for four different statements about doing evaluation: 1) I regularly collect data on my programs; 2) I regularly use the data I collect to inform programming decisions; 3) When designing or refining programs (large or small), one of the first things that I think about are the intended outcomes of the program; and 4) When considering what data to collect about my programs, one of the first things I think about is how best to measure progress toward the intended outcomes. Participants then rated themselves using the stages of change: disinterested, deliberating, designing, doing, and deepening.

Participants were asked about specific evaluation practices they currently used (e.g., administer surveys, conduct quantitative analysis). They also responded to questions about their current attitudes toward evaluation and their hopes or concerns for future engagement with evaluation.

² The organization-wide context was still in progress at the time of writing. More data collection was scheduled to be completed at the end of the ECB effort.

Formative Evaluation

In the two-year, department-wide ECB process, formative data collection occurred post-trainings and after the completion of the first year. Department leadership shared a questionnaire with staff and facilitated a group conversation. The questions focused on personal growth in understanding and implementation of evaluation, continued challenges in understanding and implementing evaluation, identifying useful resources and training, and planning for future support.

In the network-wide context, at the midpoint of the process, participants completed a modified survey based on the original needs assessment survey to monitor progress and identify potential areas to focus on or revisit in future discussions and coaching.

Summative Evaluation

Both the department-wide and network-wide contexts had summative data collection as well. The department-wide participants engaged in a reflection activity and discussion including questions about how changes in their evaluation practice influenced their teaching, how their evaluation practice impacted program participants or visitors, and new skills and knowledge learned through the ECB effort.

In the network-wide context, participants took the modified survey one last time at the end of the project. In addition, each participant took part in a semi-structured interview. Given the dispersed nature of the network-wide effort, the ECB facilitators felt that individual interviews would be more focused and insightful than doing a single group interview. The questions covered reflections on the capacity building process, individual professional growth, organizational growth, and collective capacity in their community.

Facilitator Reflection

Throughout each of the ECB efforts, the facilitators routinely debriefed and reflected on how training, coaching, and evaluation projects were progressing. After each training, the facilitators documented successes and challenges of activities and content. The facilitators also documented all meetings and conversations with department/organization leadership and ECB champions.

Findings

While it was not the original intent to compare the three different contexts, data and facilitator reflections provided insight to strengths and weaknesses of each implementation, which has subsequently informed a proposed model for the future. What follows is a brief synthesis of the themes that emerged from analysis of data collected in each ECB context, triangulated with facilitator reflections. While the organization-wide ECB effort was still underway at the time of writing, findings from the organization-wide context are shared when possible.

Evaluation Skills and Confidence

Participants across all three contexts demonstrated personal growth in their evaluation literacy. In the network-wide context, open-ended survey responses and interviews demonstrated these participants experienced a positive shift in their confidence and attitude toward conducting evaluations. For example, one participant stated:

I am confident in my ability to evaluate our programs and believe evaluation is an important and necessary step. I've discovered that I enjoy working on evaluations.

Participants also described improved understanding of how to do evaluation, including use of a diverse range of methods:

I think my biggest growth has been around the exciting revelation of the many ways to

implement an evaluation of a program--there are so many tools and methods that can be

used, and working to really understand your goals in order to choose the best method of

data collection. That has been a game-changer for me in the realm of evaluation.

Comparisons of survey results from pre-capacity building effort to post showed improvement in participants' willingness and readiness to engage in evaluation activities. There was a positive shift in participants who self-reported "doing" data collection regularly from pre (n=4) to post (n=5). Prior to participating in ECB, one of the participants self-reported being "disinterested" in thinking about outcomes first when designing programs; in their post report, this shifted to "doing." Likewise, in pre-survey responses, half of the participants (n=3) were "disinterested" in thinking about measurement of outcomes; post-participation this dropped to none.

In the department-wide context, group interviews revealed staff improved their confidence to do evaluation. The head of the department stated, "[Staff have] a greater sense that they can do this, but also a recognition that it isn't something to be addressed casually or lightly." Another member of the leadership team in the department shared that there had been "lots of revelations around the logistics of collecting data. It is not as easy as people thought, particularly with observations." Even with recognition of the complexity of evaluations, staff were excited to move forward with integrating it into their practice. A staff person stated there is a "willingness to try new things. Lots of eagerness around trying the questions in the question bank, retrofitting existing surveys, and observing participants/visitors."

Improved Program Planning and Instruction

Findings demonstrated participants in the ECB efforts became more intentional and data driven in their program planning and instruction. In the department-wide context, all participants created logic models for programs. Staff said they were more aware of how their messaging and programs related to their intended outcomes, they made changes in volunteer training to align with their outcomes, and they were more consistent in how programs and messaging were crafted. Staff expressed pride in having evidence to support what, why, and how they implemented programs or made changes.

Staff shared examples of how they planned to use the visitor or program participant voice to inform programs. Many staff had never engaged in audience research or front-end evaluation prior to the capacity building effort. They now understood the value of taking the time to hear what their audience had to say about a learning experience. One of the evaluation projects for this group focused on understanding K-12 teachers' interests and motivations for engaging in the zoo's outreach programs. Data revealed insights that the staff had not anticipated uncovering and subsequently informed future plans.

Likewise, in the network-wide context, participants reflected on how they planned to use evaluation to inform decisions moving forward. One participant stated:

Evaluation has always been a part of [our organization]. But it's a lot of numbers, quick outputs. Those don't help you create a better program.... I really enjoy how evaluation can lead us to the programs we want to be giving, instead of just checking boxes with evaluation.

Overall, participants came to see evaluation as a tool to inform their practice and create an effective learning experience for their audiences.

Culture of Evaluation

In the department-wide context, findings showed that the department adopted a culture of evaluation. Staff had a shared vocabulary to talk about evaluation. They became accustomed to a cycle of feedback and using data on a regular basis. The department created and used tools such as a question bank for surveys and interviews, and an evaluation planning template with accompanying criteria for prioritizing evaluation activities. The staff went from disparate experiences and attitudes toward evaluation to a common, unifying starting point from which to engage in evaluation practices. The head of the department noted:

[There is a change in the] conversation around the office. They use terminology they didn't have before, and now I hear it. Another evidence of this change is that other teams are asking us to be a part of the evaluation they are doing. The marketing team is rethinking the zoo maps, and it is the first time they're going to ask guests about this. They asked the education team about how to do this and to use their systems.

The department also created an evaluation intern position after the second year of capacity building to help continue evaluation efforts during peak programming and visitation periods.

In the network-wide context, there was less evidence of a culture of evaluation in part due to the fact that the capacity building effort engaged only one representative of each participating organization. This also created a challenge in effecting systemic change as the knowledge and skill building resided solely with an individual. However, there was evidence that the ECB effort helped instill a common set of evaluation values across the collaborating organizations. In final interviews, one participant commented about the benefit of engaging in this capacity building with representatives from other organizations:

One of the benefits of this was working with partner organizations to develop a shared understanding and value for evaluation. It helped us work past evaluation as a measure of outputs to looking deeper into programs from a goals-based approach.

While their home organizations had variable cultures of evaluation, when these representatives came together to collaborate on programs and other regional efforts, they now had a shared understanding and appreciation for including evaluation in their work.

Sustaining Change

Lastly, participants in each context expressed a desire to sustain and grow their evaluation practice, but the degree to which this was accomplished was dependent on a number of factors.

Of the three contexts, the department-wide ECB effort demonstrated the most success in sustaining change in evaluation practice. Staff in the department continued to seek technical support and coaching from the ECB facilitators long after training and evaluation projects were complete. Leadership shared stories of evaluative thinking surfacing in day-to-day conversations about programs and audiences. The department continued to apply a cycle of evaluation planning to identify and prioritize programs that would undergo evaluation each year. Lastly, other departments in the organization referenced evaluation work from the education department and began to incorporate similar processes for other efforts at the zoo.

Taylor-Powell and Boyd's (2008) framework for ECB references a number of the factors that may have influenced the department's success in sustaining change. First, the staff engaged in a series of professional development opportunities tailored to learning about and applying evaluation in the informal learning context. Over the course of two years, staff engaged in training, evaluation projects, coaching, and technical assistance, all of which fulfill Taylor-Powell and Boyd's framework for professional development. Second, the department secured facilitators with not just evaluation expertise, but ECB expertise. Third, they allocated time and financial resources toward the effort and utilized technology to provide access to evaluation resources. The facilitators created an evaluation resource hub with resources specific to their context. Lastly, the department had a prime organizational environment to sustain change, including a champion in the department to shepherd the evaluation efforts forward, buy-in from leadership, and established systems such as policies and standardized tools for use in the future.

It is less evident what change may have been sustained in the network-wide and organization-wide contexts. In the network-wide context, any sustained

change would be at the individual level as it was only one representative from a larger organization who engaged in the capacity building. Interview responses from two participants suggested sustained changes in their mindset about evaluation and increased interest and intention to do evaluation. One participant came from an organization that also fulfilled many of the organizational environment factors from Taylor-Powell and Boyd's framework: there was buy-in from leadership, an evaluation champion, and allocation of time and resources.

An ECB Model for Informal Science Educators

Review of the findings from data collection in each context coupled with the authors' firsthand observations and reflection on the ECB efforts revealed salient features of an ECB process to inform a model for building informal science educators' evaluation literacy.

While Preskill and Boyle's (2008) and Taylor-Powell and Boyd's (2008) models for ECB are comprehensive, they are also complex and may be difficult for practitioners to understand and navigate. In addition, they do not provide a recommended sequence of activities. Each component part of their models is important, yet comparison of the three implementations of ECB in this article has shown that some components are perhaps more integral to successfully developing informal science educators' evaluation literacy and subsequent practice.

The following proposed model (Figure 2) integrates aspects of Preskill and Boyle's and Taylor-Powell and Boyd's models but provides a recommended sequence and highlights contextual factors influencing the activities. First and foremost in the model is a needs assessment. This step is crucial to identify the motivations, needs, and current state of evaluation practice for those who will be involved in the ECB effort. This ensures that the activities of the ECB effort are best matched to the needs of those involved. For example, in the network-wide context, through the needs assessment the facilitators discovered that the participants were already actively engaged in logic model development and did not want to allocate additional time and resources toward this skill set. This allowed the facilitators to prioritize other evaluation skills in the training and coaching activities.

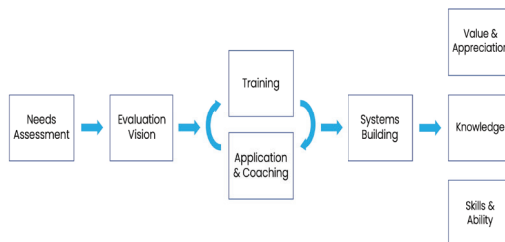


Figure 2. A Model for Developing Evaluation Literacy for Informal Science Educators

The needs assessment also influences the next component of the model: creating a vision for evaluation. This activity was fruitful across all three ECB contexts. Creating a vision for evaluation engaged stakeholders in discussions about their end goal. Just as with any comprehensive evaluation process, this allowed the stakeholders to start with the end in mind. The vision statement also served as a keystone, a reference to return to along the way to monitor the progress of the ECB effort. The vision statement also informed the design of the training and coaching activities. For example, in the department-wide context, the vision statement acknowledged the role of everyone in the department in supporting and facilitating evaluation. This highlighted the need to create a baseline of knowledge and skills for all staff in the department. In contrast, the vision statement for the organization-wide effort recognized the need to value and support evaluation at the highest level but did not prioritize having every staff person involved in the process. This helped the facilitators pivot their approach to more of a top-down model, working first and foremost with the leadership team and then bringing on specific staff as appropriate to their role and the organizational need. Both approaches had successes and challenges, and decisions about managing ECB processes from the top-down vs. bottom-up will ultimately depend on organizational culture and norms.

Next in the model is a recurring cycle of training, application, and coaching. These activities were mainstays of all three ECB efforts and proved to be the most successful in developing evaluation literacy. However, both participant and facilitator reflections identified the importance of grounding the entire ECB effort in the context of informal learning, rather than just learning about evaluation in a broader sense. In practice, this means participants actively doing evaluation projects focused on the programs and audiences they serve. It also means including case studies and examples from other informal learning evaluations to demonstrate how evaluation works in this context. When training on data collection methods, particular attention is paid toward non-traditional methods, moving beyond surveys and tests as these oftentimes do not function well with the more fluid nature of informal learning. For example, teachers often bring school groups to informal science learning settings to learn outside of the classroom and get an authentic, hands-on learning experience. Assessing students in the same manner as they are assessed in school would be contrary to the goals of an informal learning experience. Informal science educators should be challenged to assess learning and evaluate program effectiveness in a seamless, integrated manner as to not interrupt the unique learning experience.

Following training and coaching in the model is the development of evaluation systems. These systems include protocols, policies, instruments, report templates, and more. This is comparable to the structures, policies, and procedures in Taylor-Powell and Boyd's model. However, in this model the training and application of knowledge and skills through evaluation projects informs what is developed and subsequently cycles back to continued use, reflection, and further training and practice. In the department-wide ECB context, this cycle surfaced through the creation of question and embedded assessment banks (repositories of standardized questions or activities to use for data collection related to their conservation education outcomes). These resource banks were then put into practice and further refined through reflection on implementation.

The ultimate purpose of the proposed ECB model is to develop informal sci-

ence educators' evaluation literacy. This entails valuing and appreciating the role of evaluation in their practice, understanding evaluation processes and practices, and having the skills and abilities to engage in evaluation. Achieving this would complete the core competencies informal science educators need to be effective in their practice.

Discussion and Conclusions

The proposed model for developing informal science educators' evaluation literacy reflects a number of important factors to consider when embarking on ECB. First, it is important to recognize that ECB is a process and begins with identifying the particular needs of an individual, team, or organization. Not everyone begins an ECB process with the same starting point. Similarly, not everyone will have the same vision for the evaluation culture they are working toward. Articulating an evaluation vision statement early on will ensure the ECB process is relevant and effective.

Second, learning about and applying evaluation will resonate with participants when it is embedded in the context of an informal learning environment. The training, application, and coaching cycle should be specific to informal science learning. This includes using examples or case studies from informal science, adapting methods to suit informal science settings, and practicing with real-world informal science programs or experiences.

Third, ECB should include the development of systems to ensure informal science educators have the tools to sustain change in their practice. This also supports the cultivation of an evaluation culture in a team or organization. These systems may consist of instruments, policies, or other norms essential to implementing evaluation. As these systems are developed, they should subsequently be adopted through continued training and application.

Fourth, one of the challenges for any informal science learning organization is attrition of staff and volunteers. Turnover is a natural occurrence but leaves a potential evaluation literacy gap. The ECB effort should take this into consideration and create resource banks and other opportunities to ensure carryover with new staff and volunteers. Organizations should consider integrating evaluation training into the onboarding process and perhaps even match a current staff person as an evaluation mentor for new staff and volunteers.

Lastly, as Taylor-Powell and Boyd highlighted, it is important to have a facilitator with expertise in evaluation *and* capacity building to help coach individuals engaged in the effort. Effective ECB is only as good as the facilitators implementing the process and assisting the participants. Furthermore, informal science educators operate in a unique learning setting. Unlike formal education, informal educators work with a wide range of audiences, instruct or interpret in a variety of settings, and deliver programs ranging from brief encounters to multi-day intensive experiences. An ECB effort must account for this diversity of learning contexts and provide practical experience conducting evaluation in them. The authors' experience of facilitating ECB in three different informal learning contexts also suggests that, just as Taylor-Powell and Boyd recommend an ECB facilitator be knowledgeable in evaluation and capacity building, for informal science educator

ECB, the facilitator should also be knowledgeable in informal learning contexts.

Evaluation literacy is an overlooked but crucial competency for informal science educators. While evaluation is gaining more attention in informal science learning, the field is still woefully lacking adequate opportunities to develop individual, departmental, or organizational evaluation capacity. Individual educators may engage in disparate workshops or seek out resources online. However, just as when learning scientific content, a one-and-done experience is insufficient. This retrospective synthesis of ECB implementation in three different informal science learning contexts brings focus to the mechanisms needed to develop informal science educators' evaluation literacy. It helps address the gap in professional development for evaluation that is so prominently missing from the professionalization of informal science educators.

References

- Allen, L. B., & Crowley, K. J. (2014). Challenging beliefs, practices, and content: How museum educators change. *Science Education*, 98(1), 84-105.
- Ash, D.B., Lombana, J., & Alcala, L. (2012). Changing practice, changing identities as museum educators. In E. Davidsson, A. Jacobson (Eds.), *Understanding interactions at science centers and museums: Approaching sociocultural perspectives* (pp. 23-44). Sense Publishers.
- Busch, Tate-Stevenson, Green & Chesnut (2019). *Environmental education in North Carolina: examining North Carolina's environmental education providers as a community of practice using social network analysis*. North Carolina State University.
- Busch, K.C. (2020). Pedagogical content knowledge for informal science educators: Development of the iPCK framework. In review.
- Carleton-Hug, A., & Hug, J. W. (2010). Challenges and opportunities for evaluating environmental education programs. *Evaluation and Program Planning*, 33(2), 159-164.
- Castle, C.M. (2006). Blending pedagogy and content: A new curriculum for museum teachers. *Journal of Museum Education*, 31(2), 123-132.
- Ellenbogen, K. (2014). Summary of the CAISE convening on building capacity for evaluation in informal science, technology, engineering and math (STEM) education.
- 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.
- Hunter, N., Storksdieck, M., & Staus, N. (2018). Validating an Informal science learning (ISL) framework: Findings from empirical research. Technical Report. Oregon State University.
- Khalil, K., Ardoin, N. M., & Wojcik, D. (2017). Social learning within a community of practice: Investigating interactions about evaluation among zoo education professionals. *Evaluation and Program Planning*, 61, 45-54.
- King, H., Steiner, K., Hobson, M., Robinson, A., & Clipson, H. (2015). Highlighting the value of evidence-based evaluation: pushing back on demands for 'impact'. *Journal of Science Communication*, 14(2), A02.
- Kubarek, J. (2015). Building staff capacity to evaluate in museum education. *Journal of Museum Education*, 40(1), 8-12.
- Kubarek, J., & Trainer, L. (2015). Empowering museum educators to evaluate. *Journal of Museum Education*, 40(1), 3-7.
- Kubarek J. (2017) The need and timeliness of preparing informal science educators. In P. G. Patrick (Ed.), *Preparing informal science educators*. Springer.
- 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*, 1-32.
- Luebke, J. F., & Grajal, A. (2011). Assessing mission-related learning outcomes at zoos and aquaria: prevalence, barriers, and needs. *Visitor Studies*, 14(2), 195-208.
- National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. The National Academies Press.
- Patrick, P. G. (2017). *Preparing informal science educators*. Springer.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of*

A Model for the Development of Evaluation Literacy

Evaluation, 29(4), 443-459.

Prochaska, J. O., Johnson, S., & Lee, P. (2009). *The transtheoretical model of behavior change*.

Taylor-Powell, E., & Boyd, H. 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*.

Tran, L. (2007). Teaching science in museums: The pedagogy and goals of museum educators. *Science Education*, 91(2), 278-297.

Tran, L., & King, H. (2007). The professionalization of museum educators: The case in science museums. *Museum Management and Curatorship*, 22(2), 131-149.

Tran, L. & Halverson, C. *Reflecting on practice*. Lawrence Hall of Science. Accessed on July 23, 2020: <http://reflectingonpractice.org/>

Ucko, D. A. (2010). *NSF influence on the field of informal science education*. In Annual Conference of the National Association for Research in Science Teaching (NARST), Philadelphia, PA.

Zint, M. (2020). My environmental education evaluation resource assistant. Accessed July 14th: <http://meera.snre.umich.edu/>