



center for advancement of
informal science education

Summary of the CAISE Convening on Building Capacity for Evaluation in Informal Science, Technology, Engineering and Math (STEM) Education

By: Kirsten Ellenbogen
President and CEO, Great Lakes Science Center
CAISE Co-Principal Investigator

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). There are many complexities inherent in evaluating free choice informal STEM learning settings and experiences. Evaluators working in these environments address the complexities by drawing upon many different disciplines, including developmental psychology, classroom-based assessments, and health education evaluation. Yet challenges remain and are perhaps growing in this era of increasing accountability. This multi-disciplinary, maturing community needs resources to improve practice and to better support its work and the goals of informal STEM education.

This paper summarizes a Center for Advancement of Informal Science Education (CAISE) convening (June 20-21, 2013) designed to facilitate discussion about the resources needed to improve the quality of evaluation in ISE. Participants included evaluators currently practicing in the field, as well as those working in other disciplines; learning researchers; experience and setting designers; organizational leaders; program officers from the National Science Foundation (NSF); other federal funding agencies; and private philanthropic foundations. Some of the context-setting for the convening included research and development frameworks recently introduced at the federal level including the National Research Council's [Science, Technology and Innovation Indicators](#) and a preview of the since-released [Common Guidelines for Education Research and Development](#) developed by the National Science Foundation and the U.S. Department of Education. In a pre-meeting online [forum](#) participants from the larger community began to identify critical needs in the practice of ISE evaluation and made suggestions for resources, training, and other supports to advance the profession. By overall design, however, this convening raised critical questions for the field, rather than making definitive recommendations.

Three dominant themes emerged during the convening: (1) **Shared use of evaluation measures and aggregation of findings**; (2) **Access to and coordination of resources**; and (3) **Professional development**. Five additional,

Suggested Citation:

Ellenbogen, K. (2014). Summary of the CAISE Convening on Building Capacity for Evaluation in Informal Science, Technology, Engineering and Math (STEM) Education. Washington, DC: Center for Advancement of Informal Science Education (CAISE). http://informalscience.org/research/ic-000-000-010-034/ECB_Convening_Summary

less urgent topics were identified: Advocacy for the Value of Evaluation; Evaluation as a Learning Process; Focus on Science Learning; IRBs; and Broadening ISE's view to Include Other Disciplines. The notes below detail the discussions around the three primary themes. All of the convening materials, including a participant list, are available on the CAISE InformalScience.org website in a Group forum: <http://informalscience.org/community/groups/details?id=21>. Log into and/or join InformalScience.org to access all of the documents.

Shared Measures and Aggregated Findings

This theme of shared measures and aggregated findings is a frequent but controversial refrain in ISE evaluation, i.e. are there a set of outcomes and indicators that would be informative and useful to the ISE field as a whole? Outcomes that are highly specific to an individual project make it difficult to compare with others and generate new ideas. If more projects and organizations use a common framework of measurements we would have a larger, more coherent evidence base to support the value of ISE. If field-wide outcomes, indicators, and/or measures are not practical, at what grain-size would they be useful or possible? What are some examples that are being developed or tested? There are many stakeholders to consider, including NSF-funded principal investigators and project leaders, evaluators, and learning researchers, as well as program officers from public and private foundations, leaders of federal agencies, and elected officials.

Given the range of interests and concerns of those who participated in the convening discussion, it became clear that the dominant question was not *how* to develop shared measures and aggregated findings, but instead to ask, should doing so be the highest priority for the field at this time? Participants suggested that the first priority should be to identify the field's shared *goals* so that we can determine if shared measured and aggregated findings will support those goals. Hence, among the questions raised at the convening were: Is the aim of ISE evaluation to demonstrate the benefits of ISE? To compare the impact of different fields of education? To meet national standards of measurement? To improve practice? To justify funds invested in ISE? While grappling with these questions, convening participants agreed that the field does need shared measures and aggregated data, and that a role that CAISE could play would be to convene and connect the community to continue to lay the groundwork for a coordinated understanding about the purpose(s) of ISE evaluation.

There was also consensus that the informal STEM education field needs different ways of understanding its impacts. Theoretically it should be possible to make comparisons across programs. It is also important to note that the notion of common measures in this discussion does not mean singular measures. There was not an expectation for a one-size-fits-all evaluation tool. Instead, participants began to explore the issue of what tools are or would be useful.

Participants agreed that informal STEM education includes many different contexts, participants, and practices, so that comparisons of outcomes across projects may be difficult. However, “subsets” of informal STEM education types might be well served by common measures. One example discussed at the convening came from the citizen science sector, where the [Developing, Validating, and Implementing Situated Evaluation Instruments \(DEVISE\)](#) project is developing definitions of constructs that can be measured across citizen science projects. Some, such as interest in and efficacy for science, could probably be applied to other ISE settings. Other DEVISE constructs, however, such as perception of science, may not be appropriate across all STEM areas—because the lens that participants bring to engaging in citizen science activities such as birding or star watching could be different than the lens of individuals interacting with a science center exhibit on avian or astronomical phenomena.

Hence, if there is not “One ISE” then how can there be one measure? One of the signature characteristics of the many practices that fall under ISE is differentiation. “Common measures” could refer not to a single measure, but rather a set of multiple, connected measures with clear definitions. Acknowledging that the rationale required further discussion, participants identified the activities and key resources that would be needed to lead to common measures as (1) continuing to gather and organize an evidence database on InformalScience.org; (2) increasing the quality of smaller, highly-project-specific evaluations; (3) conducting larger scale evaluations across significant segments of the field; (4) identifying shared outcomes that provide in-depth understanding across projects; (5) developing meta analyses of existing evaluations. Participants recommended that CAISE take a leading role in convening and connecting those already involved in these types of activities as well as those interested in contributing further to explore collaborations and set an agenda for the future.

Post-Convening Follow-Up: CAISE collaborated with SRI International to convene a December 2013 working meeting focused on mapping assessments that can be used across learning environments. Researchers and practitioners from six projects participated in the convening. For a brief overview of the meeting visit the InformalScience.org *Perspectives* blog at <http://informalscience.org/perspectives/blog/updates-from-the-field-meeting-on-assessment-in-informal-science-education>. CAISE invited a representative group of these participants to facilitate a larger discussion on measuring learning across ISE projects in a breakout session at the Advancing Informal STEM Learning (AISL) Principal Investigator meeting in August 2014. A blog post that summarizes the major points addressed at the session can also be found on the InformalScience.org *Perspectives* blog at <http://informalscience.org/perspectives/blog/measuring-learning-across-ise-projects>.

Access to and Coordination of Resources

Informal STEM education professionals have rarely had easy access to the learning research and other resources required to stay up to date on current research and practice. This issue is exacerbated by the strong representation of unpublished “grey literature” reports in our field. The convening discussion homed in on CAISE's InformalScience.org website as a sound institutional infrastructure for strengthening and advancing the community's access to this literature. Informalscience.org is an existing aggregation tool that could continue to expand with an eye toward an enhanced social system and community activity that leverages crowdsourced materials from the wide range of disciplines in ISE. There was also dialogue about other sources of potentially useful aggregated data, including that gathered by the [Online Project Monitoring System \(OPMS\)](#) as resources to inform and strengthen ISE evaluation.

A key part of the discussion about access to resources was crowdsourcing. For example, the notion of developing an active group that produces information – through polling members – about what evaluation resources are most needed or useful. An increased community presence focused on sharing resources could be built up through a monthly online conversation that would highlight a select group of important resources on one topic. Crowdsourcing could also identify critical topics for synthesis documents for practitioners and perhaps even different syntheses for different audiences.

One approach to accelerate the process of sharing materials might be a memorandum of understanding that organizations and funders would sign to commit to sharing evaluation reports and even instruments. In addition, funders could put a stronger emphasis on requiring or recommending that recipients of their funding share their evaluation reports, just as the National Science Foundation's Advancing Informal STEM Learning (AISL) program requires principal investigators to post summative evaluations on InformalScience.org. Perhaps most critical is the need for easy, online, affordable access to peer reviewed journals via the publishing companies. An even more ambitious strategy would be the establishment of a section of InformalScience.org that allows the uploading and sharing of raw data.

Post-Convening Follow-Up: CAISE continues to collect and curate syntheses of relevant research to be added into the ISE Evidence Wiki (<http://informalscience.org/research/wiki>). InformalScience.org now also offers users *EBSCO Education Research Complete*, an online research database that includes full-text access to articles and papers from over 1,000 journals and abstracts from more than 2,400 others. By enrolling as a member of site, ISE professionals can log in to the [EBSCO and Journals](#) page to search and download research articles.

Professional Development

The need for professional development in evaluation extends to both ISE evaluators and those working with them. The major issues here can be summarized as: rigor, epistemology, approaches, cultural competence, and effective use of evaluation results.

There is also an urgent need for systematic professional development trajectories for newer evaluators of ISE (some guidelines for this have already been created by [Visitor Studies Association \(VSA\)](#) and [American Evaluation Association \(AEA\)](#). As professional development opportunities are created to address the identified trajectories, it is important to ensure that they are high quality, affordable, accessible, and culturally competent. (A formal certification process was not a primary focus in the discussion at the convening.) Participants identified multiple organizations that could lead this effort: The Afterschool Alliance, the Association of Science-Technology Centers (ASTC), the American Alliance of Museums (AAM), Committee on Audience Research and Evaluation (CARE), and the Visitor Studies Association (VSA).

With respect to professional development, the discussion group identified three types of evaluation professionals:

1. **Evaluation “technicians”:** These professionals are often accidental evaluators who have been asked to add evaluation to their job scope. They tend to need tools rather than a comprehensive training.
2. **Evaluation practitioners:** These professionals may have specific evaluation training but need ongoing updates to hone their skills as well as keep current on emergent developments and new concepts.
3. **Evaluation champion administrators:** This is the principal investigator or organizational leader who hires and works with evaluators. They may need training and resources for working with evaluators and for developing project outcomes.

Post-Convening Follow-Up: CAISE has created a newly formatted and downloadable PDF version of the [Principal Investigator’s Guide to Managing Evaluation in Informal STEM Education Projects](#). This resource can be a particularly critical tool for the “evaluation champion administrator” type of stakeholder identified above. Ongoing development work on the organization of the [Evaluation pages](#) of InformalScience.org continues and is informed by discussions and outputs from this convening as well as by formative evaluation findings.