ISEE Strategic Outcomes Framework for Measuring Informal Education Outcomes and Institutional Impact:

Revisions and Additions in Response to Expert Explorations April 2021

Compiled by
Deborah Wasserman, Ph.D., COSI CRE
Colleen Popson, NMNH

April –November, 2021

This material is based upon work supported by the National Science Foundation under Grant No. (NSF AISL #2039209). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

Table of Contents

The Informal STEM Education Evaluation System

In April 2021, the Center for Evaluation and Research at the Center of Science and Industry (COSI) with the National Museum of Natural History (NMNH) and Partnerships in Education and Resilience (PEAR), hosted a National Science Foundation-funded virtual conference aimed to contribute to the development of an innovative informal STEM education planning and evaluation system for Informal STEM Education institutions. The Strategic Outcome Progressions Conference: Exploring a Framework for Measuring Informal Education Outcomes and Institutional Impact (NSF AISL #2039209) convened a small, diverse group of 24 research, evaluation, and education experts to face validate an outcomes framework that informal educators can use for strategic planning and evaluation of programs that aim to broaden participation in science. The conference and the framework that was its subject are part of a larger initiative to develop an Informal STEM Education Evaluation (ISEE) system that could be used by any informal science institution to support strategic development, implementation, measurement, and results reporting of STEM education plans. Unique to this system is the ability to unify program and exhibition planning and evaluation across an entire institution.

The ISEE System is an idea being researched and developed by Smithsonian's National Museum of Natural History and the Center of Science and Industry's Center for Research and Evaluation, with guidance from Partnerships in Education and Resilience (PEAR). NMNH and COSI have been incubating the system through several rounds of pilot projects within their institutions.

The intent of the system is to address challenges shared by informal STEM education institutions in the strategic planning, design, implementation, evaluation, and reporting of their education efforts. By nature of the broad-ranging topic areas they address and the free-choice settings in which they take place, informal educational programs can vary significantly even within a single learning institution, in content focus and learning goals; format depth, duration, and dosage; and audience age, cultural background, interest, and composition. This variance makes it difficult for educators to target appropriate learning outcomes for strategic planning, program design, and evaluation--and even harder to track an institution's impact across all these varied purposes and audiences.

To date, ISE researchers and educators have addressed these problems in innovative and important ways. They have enumerated informal education outcome areas, especially as they differ from formal education (Bell et al., 2009; Friedman, 2008); created guidelines for programming with attention to diversity, equity, access, and inclusion (Garibay & Teasdale, 2019; Urban, Archibald, et al., 2021); identified domains for evaluation measurement (Teasdale, 2021); and created shared measures so that similar concepts can be measured similarly across programs and institutions (Sneider & Noam, 2019). Moreover, evaluation researchers have also created systems for integrating evaluation into planning. Innovation Network's Point K Learning Center (Innovation Network, 2012) provides online mapping for building program logic models through to evaluation. The Evaluation Netway (The Netway, 2021) additionally nests a sophisticated online process for identifying and measuring program outcomes within a "partnership model" and in line with its evolutionary stage.

These tools have significantly advanced planning and measurement of informal education programming (Allen & Peterman, 2019). However, across departments within a single institution (e.g., exhibitions, outreach, and education) and even programs within individual departments, these tools are most often utilized differently and independently often with highly varied definitions and measures. Thus, an institution's educational impact is extremely difficult to both plan and measure. Allen & Peterman (2019) also challenged the field to find new methods, specifically ones for situating programming within its broader ecological context. The ISEE system meets that challenge in two ways. First, it helps an institution see beyond its individual programs to frame, measure and understand the ways its programming

portfolio contributes to a larger ecological picture. Second, particularly through its strategic outcomes framework, ISEE provides a way for educators—and learners themselves—to chart a learning course, creating opportunity both within and beyond a single institution. As a whole, the system combines these new innovations with multiple aspects of existing tools in a way that unifies, within a single institution, the planning and measurement of the wide and varied programming that, when considered in its entirety, meets the institution's educational mission.

The system consists of three fundamental components:

1) a framework for strategic learning outcomes (currently in the validation phase),

2) a framework for strategic diversity, equity, inclusion and access (DEIA) integration (currently in the ideation phase), and

3) an evaluation platform (currently in a proof-of-concept phase) that allows educators to create customized surveys based on shared, agreed-upon measures and that displays aggregated and filterable survey results to provide at-a-glance views of how an institution's programs are affecting their audiences in various areas.

The first component – the framework for strategic learning outcomes, the subject of this white paper, has received the most attention in our pilot work at COSI and NMNH. This framework builds on previously defined informal learning categories and strands (Bell et al., 2009; Friedman, 2008) by delineating outcome expectations based on dosage and complexity of programming. Although central to the ISEE system, in its own right this innovation addresses Allen and Peterman's (2019) challenge to "design methods that characterize the impact of a STEM resource or program not just as a stand-alone offering, but in terms of its capacity to support and connect to other experiences, resources, and programs in the ecosystem" (p 29-30).

More specifically, the outcomes exist within eight categories (interest, attitude, knowledge, STEM skills, 21st century skills, behavior, science capital, and career activation and preparation). Each category includes roughly six outcome types that can be mapped to programming that ranges from short duration with simple content (e.g., a hallway cart demonstration can trigger situational interest) to extended duration with complex content (e.g., a youth internship can generate well-developed personal interest). Each outcome type is then specified to be relevant to a given program topic and its intended learners. The framework has the potential to provide informal education institutions a tool for both program-level and institution-level strategic educational planning and evaluation. Exploring the utility of the framework was the subject of the NSF-funded Strategic Outcome Progressions Conference in April 2021.

Along with a brief conference description, this document summarizes the themes raised and our responses and changes made as a result. The themes documented here derive from analysis of dozens of pages of conference Moodle discussion notes, chat notes, and session notes. The complete analysis is available upon request.

Conference Description

The Strategic Outcome Progressions conference convened 24 experts from around the United States to explore the face validity, utility and feasibility of a framework for strategically planning learning outcomes and assessing impact in informal STEM institutions. Participants, their affiliation, and their conference role can be found in Appendix B. Conference participants reviewed and contributed to the soundness and utility of the outcome categorizations as well as the outcome types within them. These insights and inputs have furthered the development of the framework toward the goal of its use as a valid and reliable tool.

The conference was organized and run by COSI's Center for Research and Evaluation, Smithsonian's National Museum of Natural History, and PEAR, and was funded by NSF (AISL #2039209). Participants included eight education practitioner and eight evaluator participants. In addition, eight research consultants provided expertise specific to each of the eight outcome progressions. Central to the framework's critical review is its relevance and inclusivity across the broadest possible range of cultures and developmental age groups. It was important that the conference participants have experience implementing (in the case of educators) and evaluating (in the case of evaluators) a wide variety of programs for diverse audiences. We intentionally recruited participants from institutions that varied in terms of size, the type of science, and location. We required that all participants have experience with programs targeting populations underrepresented in science and sought representation among the participants from those populations, including Black or African American, Latinx, and American Indian or Alaska Native.

The conference was 100% virtual, with a mix of synchronous and asynchronous sessions:

April 8, 3-4:30 pm ET, Preconference Zoom Meeting
April 8-15, Asynchronous Preparation, Moodle Discussion Board, Self-paced, 1 hour
April 15, 11am-4pm ET. Conference Day 1, Zoom Meeting, with breaks
April 15-26, Asynchronous Discussion, Moodle Discussion Board, Self-Paced, 1 hour total
April 26, 11am-4pm ET, Conference Day 2, Zoom Meeting, with breaks

The conference was spread over the course of several weeks to allow time for complex ideas and thoughts to germinate and to give the leadership team opportunities to course correct in response to discussions and to participant feedback. Continuous asynchronous discussions occurred on a Moodle learning management system, punctuated by three live Zoom sessions planned at regular intervals. Live sessions including chat discussions were recorded, notes were taken in breakout sessions, and a Mural collaboration tool was also used to capture participants' questions, ideas and concerns in breakouts. Together with the Moodle discussions, these artifacts provided a wealth of participant contributions from which to draw.

The goal of the conference was to have researchers, and informal STEM evaluators and educators convene to critically review the feasibility, utility, and face validity of the ISEE Strategic Outcome Progressions framework. It was designed so that participants would leave the conference with an understanding of:

- the current and future utility of the tool for (1) institution-wide educational program planning and evaluation (2) program-specific planning and evaluation
- direction for creating a research agenda for improving the tool's utility and validity

Conference organizers (the Leadership Team) recognized that key to effectively meeting this goal was the inclusion of a wide range of perspectives and voices. To this end, they created, as part of the external evaluation, a Culturally Responsive and Equitable Evaluation and Development committee charged with contribution to monitoring and providing feedback into planning and conference implementation. Based on the summative evaluation report (see Appendix C), this objective was met.

Strengths and Issues Raised at the Conference

The conference proceedings helped us to better understand the framework strengths, gaps, and challenges to address as we progress. Some recommendations shared were adopted both during and immediately following the conference (Section 4 – Changes in Response to Conference Participant Response), others will be instituted in future iterations, and others are still to be considered (Section 5 –

Further Questions/Future Work). Readers can find an image of the framework in the section titled Framework Visual Reconfiguration on page 19.

This section begins with a listing and description of general framework strengths, uses, and evidence of need for the framework. While this acknowledgement felt supportive and is important to forward momentum, the conference's deeper value came from the challenges to the framework. These have helped us understand the work ahead to make the framework useful and usable, especially in ways that will respect the multiple dimensions of all learners, the cultures they bring with them, and the systemic barriers that inhibit learning. Thus, the bulk of this section describes major issues to be addressed in future framework iterations.

During the conference, participants provided recommendations for framework changes, many of which evolved over the course of the conference as participants arrived at deeper understanding of the contents and the framework's utility and applicability.
The concerns, recommendations, and changes described in this section come from the following sources:

- Compilation and organization of comments made throughout the conference.
- Observation of the kinds of explanations and processes that helped conference participants better understand the framework contents and utility.
- Clarifications made during the course of the conference that furthered participant understanding.
- Review and response from Martin Storksdieck who was a researcher specifically assigned to looking for supports users will need. His feedback can be found in Appendix A.

General Strengths, Uses, and Evidence of Need

In general, despite being put in the role of critic and "torture testing" the framework, participants were enthusiastic about its potential. We begin by documenting the strengths, uses, and evidence of need their comments revealed.

Through their interactions with the framework, conference participants found that it had the potential to contribute to the informal STEM education field because of its utility as a tool that:

- Parses learning experiences into outcome types that realistically reflect contact time and experience
- Provides qualitative dimensions to existing learning categories
- Focuses program planning
- Provides overview of institutional educational intentions
- Promotes dialog about program design
- Builds evaluation capacity
- Builds practitioner capacity
- Aids in strategic planning

Strategic planning and program development. Participant comments referenced the framework as providing -- for educators, evaluators, designers, and leadership—useful guidance for building a shared language and purpose for strategic planning and program development, from the earliest stages of the program planning process through data gathering and reporting.

Promotes discussion. Participants also recognized that for all users, the framework tends to generate important discussion that might otherwise go unaddressed. In the process of selecting outcomes, for example, users are prompted to debate and more explicitly name and define the type and depth of learning they hope their programs can achieve and, following, what mechanisms and time allotment a

program would need to deliver those outcomes. When the framework is used to identify outcomes deemed primary, as well as outcomes that might support those primary outcomes, participants found themselves drawn into even deeper meaningful discussion about program intent and design approach. For example, two programs might have the same outcomes, but which is primary and which is supporting will change the nature of the program.

Capacity building for educators. Conference participants were also asked to reflect on how the framework might be useful for capacity building among education practitioners (educators), evaluators, and leaders. Responses suggested the framework could give educators license to be more explicit, realistic, and nuanced in defining the outcomes for their programs. With that, they would also be challenged to think more carefully and creatively about design implications, namely the mechanisms for delivering those outcomes. In addition, some said the framework could be useful for building practitioner skills, both for program design and implementation. For example, working with the framework could help new educators understand the full (or a fuller) range of STEM learning, how outcomes might work together for different audiences to support learning, and ways programming might be designed to achieve those mixes. One participant suggested that every practitioner should be introduced to the framework as a way of understanding the full range of informal STEM learning. It could function as an introduction to informal STEM education. It was also noted how the framework could help support educator-led evaluation when resources might not otherwise be available.

Contribution to evaluation. Participants also identified and noted ways the framework could help evaluators do their work and build capacity for guiding evaluations. For example, the framework was seen as a springboard for "helping people think more broadly" about programming and how it can function, as well as a tool for supporting and guiding conversations with educators, addressing such topics as:

- What types of outcomes—intended and unintended--do you think your program is achieving?
- What outcomes do you want your programs to achieve?

In addition, they saw the use of outcome types helping to specify evaluation questions, and as such, aiding in the selection of methods and instruments. The framework standardizes outcomes at a granularity that theoretically could allow for comparison over time and across programs. Relatedly, the framework encourages common systems and measures that individual programs can access affordably. Finally, like with educators, the framework can serve as a training tool for new evaluators or evaluators less familiar with informal education.

Discussion and decision-making by leadership. Finally, participants thought the framework could help leadership by triggering discussion and decision making around strategic reflection and planning in the following ways:

- to understand organizational strengths and roles of programs in achieving its mission;
- to assess the strengths, dependencies, redundancies, and gaps in an institution's education programming portfolio across varied audiences;
- to direct changes to those portfolios to better serve their audiences; to communicate program descriptions and impact for purposes of accreditation, grants, marketing opportunities, and brand building;
- and ultimately to help leaders identify partnerships outside of institution.

To further strengthen these contributions, participants' discussion and feedback revealed important issues needing to be addressed. Each is listed and described below.

Issue #1: Responsivity to Cultural Diversity, Equity, Accessibility, Inclusiveness, Power & Voice

Despite agreement about its usefulness, participants were particularly vocal and insightful about the need to counter the framework's attempt at universality of informal STEM learning with explicit direction for learning that is responsive to cultural diversity, equity, accessibility, and inclusivity. Participants challenged each other and the conference team to situate the framework in what one participant called "the bigger picture," with another more directly stating that the framework "does not feel equity or context focused." Noting that the framework construction had not included program participant voice and no program participant voices were present at the conference, conference participants sought to find how the framework includes topics that voice would directly affect. Specifically, these topics include:

- How intended outcomes reflect participant context;
- How outcome selection reflects participant voice;
- Where outcomes targeting transformational agency and desettling Eurocentric ways of thinking belong in the framework;
- Unexpected consequences of a codified outcome structure; and
- How programs involving constructivist learning and other kinds of creative learning environments locate themselves in the framework.

Context

The framework needs to be designed in a way that acknowledges that outcomes and outcome selection must reflect participant context. Every individual participant arrives at a learning experience from a social (including community)/emotional/physical environment that can set up learning interest, ability, and motivation, which in turn affect learning outcomes. The framework needs to function as a tool that acknowledges context in all its diversity.

Participant voice

The framework also needs to acknowledge that successful program design, implementation, and evaluation all depend on how well they integrate with the people who use them, most specifically, program participants. To include participant voice, it will be important to be mindful of power dynamics so that the framework functions in a way that responds to participant voice rather than assuming it. It will be important to find a mode of including voice that recognizes the need to challenge the status quo power dynamics and assumptions that, to date, have primarily driven STEM education.

"Authority of science" and potentially exclusionary values & constructs

Many participants identified the problem of the framework being firmly situated in a "white middle-class" worldview that accepts the authority of science. As part of this problem they noted the need for evaluation to be relevant across culturally diverse participants, age groups, and learner experiences.

Even if equity were built into the framework design, "there is a threat of ISLs assuming that they would be 'done' with equity if their programs somehow met some metric, without considering implementation or underlying pedagogical stances." There was a strong caveat that without explicit directive for relevance across cultural diversity, the framework would "norm to white middle-class values." As it stands, the framework makes invisible, culturally exclusionary assumptions about definitions and roles of STEM skills, 21st century skills, career development, and others.

Another concern is perpetuating exclusionary definitions and the systems that support them. For example, one participant presented the concept of STEM career as an example saying:

> *Youth need to jump through the hoops in the formal science classrooms in order to advance along the traditional STEM pipeline. The advancement is via formal science contexts, not informal science.*

Thus, the framework needs to be nested within a system that encourages informal science programming to differently access the traditional pipeline, and perhaps even unsettle the definitions of STEM careers.

Equity and justice work needs to be continual and is something that always needs to be evaluated and addressed.

Transformational agency and desettling Euro-centric ways of thinking

In light of this concern, participants challenged the framework with inclusion of, and even insistence on, transformational agency and desettling Eurocentric ways of thinking. "How," one person asked, "does the framework account for 'desettling' as a complex outcome?"

One aspect of this concern is the need to guard against the use of STEM as a political issue (i.e., the power it has as a "handmaiden" of other social forces). The framework, nested in assumptions derived from limited dominant-group perspectives, lends itself to perpetuation of STEM as existing within and maintaining an exclusionary power structure. How can the framework make users accountable for the diverse array of socio-cultural considerations necessitated by efforts seeking to engage the broadest of cultural diversity in STEM learning and knowledge production?

Concerns with a codified framework

There is a concern that a codified framework may provide permission to preclude necessary recurrent dialogue about such topics as: participant inclusion or exclusion, outcome relevance to participants, scaffolding based on participant assets, etc. As one participant explained:

> *If this framework becomes codified as a framework – in terms of equity and justice – [I] fear that people could check this off and say 'okay we met this goal and we are done…. Having a snapshot approach risks stagnation with equity and justice, instead of continual growth.*

A challenge to the framework is to nest it within requirements for this continual growth.

Disrupting the definition of STEM Learning and STEM Learning Categories

An important aspect of making the framework culturally responsive to and inclusive of the full range of potential learners is being open to critique of these learning categories and their intended use. Some participants challenged the group to "shake the foundation" and "disrupt the structure" asking such questions as:

- Do these recognized learning category "buckets" perpetuate "the authority of science" and the ways it has maintained inequitable power structures?
- Under what category does explicit STEM criticality belong? (e.g., questioning power and historicity)
- Does an adoption of accepted learning categories perpetuate blind spots that exclude equally as authentic or other creative ways of knowing?
- How do we challenge this model with the same questions with which we challenge the whole industry?

- Is this stagnant and old school?
- Is this "old school" way of understanding STEM learning being used as a way of wielding power? (a la Stanley Aronowitz, e.g. (Aronowitz, 2004)
- To make the framework applicable across learners, we need to know how all learners are defining and valuing not only learning, but science and STEM. Who is privy to the learning? Where in the community does STEM knowledge reside? How does science learning reflect learners' unique context, ecology, community, etc.?
- How are outcome categories such as knowledge and skill justified when pedagogy standards such as NGSS seek to de-emphasize them?

Danger of rigidity, blind spots, and discouraging constructivist learning and other kinds of creative programming

A more generalized view of the problem is that a codified listing of outcomes or even framework of how to think of them risks rigidity and exclusion not only of people, but ideas. It was generally acknowledged that the framework needs to be situated in a larger context that ensures its dynamic nature. Underlying this need was recognition of a potential tendency to make all programming fit into the framework and thereby creating "blind spots" that threaten creative and responsive programming. As part of this concern, participants recognized the often constructivist and serendipitous nature of informal learning programming. It will be important for the framework not to preclude this vital aspect of informal learning.

Issue #2: Clarity in Framework Structure and Presentation

One of the most useful aspects of the conference was what it told us about what was not working with the way we presented the framework; i.e., what we would need to do a better job of naming, defining, and scoping. First, the term "progression" was immediately flagged as problematic, as it was found to lead to various interpretive misunderstandings. In addition, it was not immediately clear how outcomes from different categories could relate to one another, or how some outcomes might be primary, while others are supporting. Finally, it was clear that there is work to do to more clearly define the outcome categories and the outcomes within those categories, including the relevant research underlying those definitions.

Throughout the conference, participants wrestled with the "order" of outcomes within each category. Although there seemed to be logical anchors at each end, the arrangement of outcomes in between those anchors was a stumbling block. The word "progression" elicited misunderstandings of the framework's intent. Misunderstandings and assumptions included the following:

- "Progression" reflected learning progression, i.e., as one participant described, "building coherence to promote productive learning."
- The framework dictates a single path by which learners accumulate knowledge within a category.
- Each learning "path" begins with a single starting point and ends with a given end point.
- Learning is linear and occurs within each of these categories.
- Outcomes within categories are inflexible and not applicable across age groups.
- Progressions need to differ based on developmental age.
- Practitioners and other stakeholders will assume that higher "level" programming is "better."
- The higher level implies educational attainment, success, or advancement, potentially suggesting that programs designed for or learning attained at the lower level outcomes are "insufficient."

Conference discussion also addressed the framework's division into eight categories, derived from NSF impact categories, which were developed to communicate to stakeholders a range of program impacts (Allen et al., 2008). Discussion about these categories focused on utility and specifically revealed the need for clear definition, distinguishing between categories, and category interdependence.

> *This discussion makes it clear that when push comes to shove, the details matter on what we mean by things like Attitudes, Interest, Career, etc. There are so many different ways to choose within each of them, and even more to operationalize and measure. And as was pointed out: they overlap. (notes from case study 1 breakout session 5)*

Discussion also addressed the division of categories into outcome types. While participants addressed outcome types both within categories and across categories, they also asserted the need for: a research base to support the outcome types and their order as well.

Conference participants also expressed concern about distinguishing between categories. Some areas that could use more clarity included:

- What are the distinctions and overlaps between interest and attitude? Between STEM skills and 21st century skills?
- What are the implications of selecting one category over another when both categories are appropriate? For example, a program that introduces participants to varied careers related to a STEM topic might identify the outcome as attitude (toward STEM careers) or STEM Career (introductions). The implications of each need to be clarified (e.g., attitude toward STEM Careers vs. STEM career introduction).
- What are the options for category and outcome-type selection when multiple categories are involved? For example, how do you account for the idea that attitudes have both an affective and a behavior component (e.g., handling snakes may involve empathy but also the skill to handling snakes properly)? Similarly, confidence building (attitude) can be built by knowledge acquisition and both can lead to consideration of a science career.

Another area requiring further clarification is that some important STEM learning outcomes (e.g., identity and self-efficacy can be defined by outcomes across multiple categories. Users need guidance specific to these often-used constructs.

Participants also noted the need for more clarification when a program may need to select a single category, but with varied specification within an outcome type as would be the case, for instance, for a program where participants learn an "ecosystem" of skills. In this case practitioners will need guidance for how to specify outcomes within outcome types.

Further, there may be the need to use the same outcome type twice but specified differently, e.g., attitude toward STEM Career; attitude toward STEM.

Issue #3: Accounting for Outcomes Not in the Framework

Although some participants saw the categories as "comprehensive" and including "all major categories of field interest," others sought to understand how they would include various types of program intentions that didn't seem to fit. Whether to include new categories to accommodate these scenarios became an important question. These included programs that address social justice issues, social emotional development and learning, STEM identity, and building STEM community and sense of belonging. Each of

these topics is described below. In addition, program intentions mentioned but with less discussion included digital co-creation, holistic approach to STEM, and self-efficacy.

Participants also noted that outcome selection can and possibly should depend on the research-based theory of change upon which the program is built. Thus, they questioned if all relevant theories of change could be translated into framework outcome types within categories.

Social justice issues and system change

A great deal of discussion acknowledged the importance and even imperative for informal science education to be addressing social justice issues. Most participants agreed that the framework needs to reflect this imperative. Included within this arena of "social issues" are racism (and anti-racism education), anti-oppression, social impact, access to and inclusion in STEM and STEM fields, voices in STEM definition and understanding, social impact on science and society, and transformational agency.

Understanding the range of how ISE programming addresses social issues provides background for determining how the framework can both accommodate, support, and even encourage this type of programming. Discussion revealed two important aspects of how STEM programming addresses social issues:

STEM and Greater than STEM Social Issues. First, ISL programs can use STEM experiences to target STEM systems specifically (e.g., access to STEM careers, inclusion of diverse STEM knowledge systems) or larger social systems (e.g. anti-racism, anti-oppression even outside of STEM systems).

Individual and Community-level Social Issue Impact. Also, ISL programs can target systemic impacts at both the individual and community level. For example, similar to a systems-level impact, "resilience" or "transformational agency" describes an individual but describes how the individual relates to multiple systems. A guide might also demonstrate how a program targets resilience as a longer-term outcome to which the program contributes and how measurable outcomes might be constructed.

Thus, within this discussion specific challenges to the framework emerged. Described below, these challenges included unit of measure, outcomes that support systems change, and strategies for assuring the framework addresses social issues.

Unit of Measure. Discussion of programming intentions involving social justice issues included consideration of the challenge that individual learning and systems change each require a different unit of analysis. Currently, the framework only includes learners as the unit of measure. Social justice evaluation would involve larger systems (e.g., communities) as the unit of measure. System-level analysis may be beyond the scope of this framework (i.e., this framework is designed to structure ISL learning outcomes). Does a community "learn"? As a longer-term impact, system-level change may be beyond the scope of this framework and probably requires contribution analysis rather than outcome measurement.

Outcomes that support systems change. The discussion of ISL institutions' commitment to addressing social justice issues leads to the need to answer the question: what types of learning outcomes contribute to social justice in larger systems? One participant, for example, noted there is currently very little language for addressing DEI outcomes for middle schoolers. Much learning that leads to social justice involves developing specific skills that could be considered "21st century skills" if that category broadens to include social-emotional development and learning as well as skills specific to anti-racism and anti-oppression. However, participants acknowledged that multiple learning steps contribute to systems change and that ISL programs seek to produce learning outcomes that may ultimately produce or at least contribute to systems change. Along with skills, these individual contributions will derive from learning

that addresses any of the remaining categories (behavior, attitudes, knowledge, interest, STEM skills, STEM capital, and STEM careers).

Some participants suggested making social change central to the framework, thus bringing into question the framework's purpose. As a unifying organizer of all ISL programming, centrality of social issues would marginalize more traditional STEM learning programs. On the other hand, the framework could challenge even traditional STEM learning programs to contribute to collective impact and social transformation.

Another challenge is to include these issues in conceptualizing framework use without making it "just a checkbox." In other words, the framework's application and structure needs to ensure that outcome selection and specification always occurs in a way that includes and respects learner voices. The process must acknowledge that, by nature of its assumptions about intended learner outcomes by program designers, outcome-based planning and evaluation risks perpetuation of social injustice.

STEM activity for greater good (e.g., community building, healing). Participants also noted that programs with intended outcomes aimed at affecting the greater good, such as community building or healing, would have difficulty locating those outcomes in the STEM learning framework. This category may differ somewhat from social justice issues because the "greater good" may not be "issue" related.

Other Commonly Used Outcomes

Social emotional development and learning. ISE programming regularly seeks to produce social-emotional development and learning outcomes such as interactions, relationships, identity of self, recognition of ability, expression of emotions, empathy, emotion regulation, impulse control, and social-emotional understanding. Participants noted that these outcomes need to be located within the framework, and perhaps by expanding the 21st century skill category to include them.

Self-Efficacy & Sense of Competence. Programming often addresses self-efficacy and sense of competence, particularly as a support toward STEM identity or simply social-emotional development.

STEM Identity. Conference participants recognized STEM identity as being integral to much ISE programming, and located it in the framework including in attitudes, STEM capital, and career path. Some suggested it be more explicit in the framework, while others suggested that by using the framework to more clearly define its purpose, it could then be located within the framework. For example, STEM identity could be an intended outcome as a way of encouraging entry into STEM careers or just a sense of "being a STEM person." Moreover STEM identity, like many intended outcomes, also needs to be defined programmatically in terms of supporting outcomes. Some programs, for example, might need to generate interest as part of STEM identity that may be an attitude toward self. Others might begin with interested participants and thus need to address attitude toward academic systems in order to support STEM identity that leads to a STEM career. Some examples of shorter-term outcomes that lead to STEM identity included recognition and performance that lead to self-efficacy and sense of competence. Diverse use of STEM identity as a primary outcome included STEM Capital (as occurs with being a community science peer leader), career choice and development, and personal STEM literacy.

Building STEM community; sense of community; or sense of belonging. Another ISE focus area that participants identified as needing a place in the framework was building a STEM community, sense of community, or sense of belonging to a STEM community. Sometimes programs exist to create a STEM community or create a sense of belonging to an existing STEM community. Still others seek to build a sense of community even if not directly related to STEM.

This sense of community can be a supporting outcome to career building.

STEM literacy, Beliefs & Values, Empathy. Participants were confused about outcome selection for programs that generally promote STEM literacy; beliefs and values, and empathy

Acquisition, Assimilation, & Transformation. In addition to outcomes that describe acquisition or assimilation, the framework outcome types need to account for transformation.

Constructivist, Serendipitous Outcomes. Many informal STEM educational experiences are teleological in nature and designed to produce serendipitous learning. Where do these outcomes fit in the framework? Perhaps, in these situations, the outcome to measure is situational interest or situational attitude toward science (e.g., science can be fun).

Issue #4: Framework Use

It will be important to clarify how the framework is to be used. As described by research consultant, Martin Storksdieck (see Appendix A), issues of purpose and target need to be clarified:

> *Is the Framework a precise guidebook for evaluation and measurement, one that allows to define exactly what to measure due to the nature of the experience that can be mapped against the Framework? Or is the Framework a planning instrument, one that allows experience designers in thinking what is possible, and to help guide decisions about goals and objectives in realistic and accomplishable ways? The meeting seemed to have pointed to the latter, given the many complexities and perceived needs for changes in the Framework. The level for needed "validity" of the Framework is much lower if it serves as a planning and design heuristic, and primarily serves to guide thinking, than if it serves as a tool to shape evaluation and research studies.*

Participants found that the framework made more sense in the context of applying it.

A participant asked, do we want to encourage non-specialists to do evaluation? Given that one purpose for the framework is evaluation capacity building, what kind of supports and guidance do practitioners need to be able to use the framework effectively?

The description of the framework needs to make clear that the ordered depth of outcome types distinguish it from other conceptualizations such as the NSF/Friedman and ISIE learning strands (Bell et al., 2009; Dierking, 2008).

Exploring the framework raised questions about how the framework or the evaluation system that applies it will accommodate the following:

- Programs occur in various "grain sizes" (e.g., a particular meeting within a topic within a series). Ideally, the framework can apply to any grain size, depending on the evaluation question to be answered.
- Varied indicators across audiences and topics. Indicator statements need to flexibly accommodate wide ranges of generic outcome type specifications and topics. If so, in and of themselves, indicator and outcome statements can, when written in a way that is responsive to program context, function to inspire programming ideas and designs—especially as their specification is allowed to differ across audiences and communities. This concern, even during the conference, led us to recognize that the outcomes and indicators listed in the framework are outcome "types" that suggest this inspirational kind of specification. Thus, specified outcome statements and indicators will differ by audiences, e.g. for families vs individuals or other groups.
- Outcome types can be used in multiple ways — all of which need to be clarified for users. These multiple uses may include:

- As coding schema for observation data
- As post-test only
- As post with retrospective pre-test
- As "reach" outcomes (i.e., how a program can set expectations that individual participants may reach different depth levels in an outcome category)

- Multiple response types, e.g. Likert-type rating? Binary yes-no?
- Additional evaluation concerns such as:
  - Outcome selection and measurement often needs to account for change from entry level.
  - Outcome evaluation needs to be nested within additional evaluation questions such as the quality of the pedagogy or differences between groups.

Changes in Response to Conference Participant Response

This section includes clarifications made since the conference and based on conference input. These clarifications have been necessary for continued piloting of the framework as it is nested within the larger ISEE (Informal STEM Education Evaluation) System.

Toward a framework that supports diversity, equity, access and inclusion

Throughout the conference participants struggled with reconciling the framework with how it can be supportive of, and responsive and conducive to issues of diversity, equity, access and inclusion at both personal and systemic levels. There were attempts and needs to infuse these concerns into each framework component: the categories, category selection, the outcome types and arrangement of types within categories, indicator statements, and unit of measure. Participant comments revealed widely varying strategies for doing so. Thus, during the conference, some working strategies emerged. Others have been developed since the conference.

During the course of the conference, we came to understand that cultural context and participant voice need to influence outcome choice and program design. This influence is seldom in and of itself the learning outcome. (Not to say it can't be; certainly programs can intend that participants find voice and set learning goals. These learning experiences can be located within and as a combination of any number of framework categories.) The Figure 1 illustrates how we came to understand the nature of cultural context and program design as program level variables that produce the outcome types found in the framework.

More specifically, we conceptualize that framework outcome types emerge from three program-level variables: the nature of a learning experience (variable #1) that derives from how learner context and opportunity (variable #2) and influences the nature of the program design (variable #3). The nature of the program experience produces learning outcomes that can be generally located in the framework of outcome types.

Program design involves: (1) outcomes specified within contexts; (2) cross-cutting (or foundational) issues (e.g., DEAI and social justice to issues of power, representation); and (3) evidence-based practices.

The framework addresses only **the outcome types that broadly describe possible learning outcomes.** Within those types, and based on program strategies, outcomes are then specified for any given program. (This conceptualization was inspired directly from an observation/comment informal STEM learning consultant Martin Storksdieck made during the conference.)

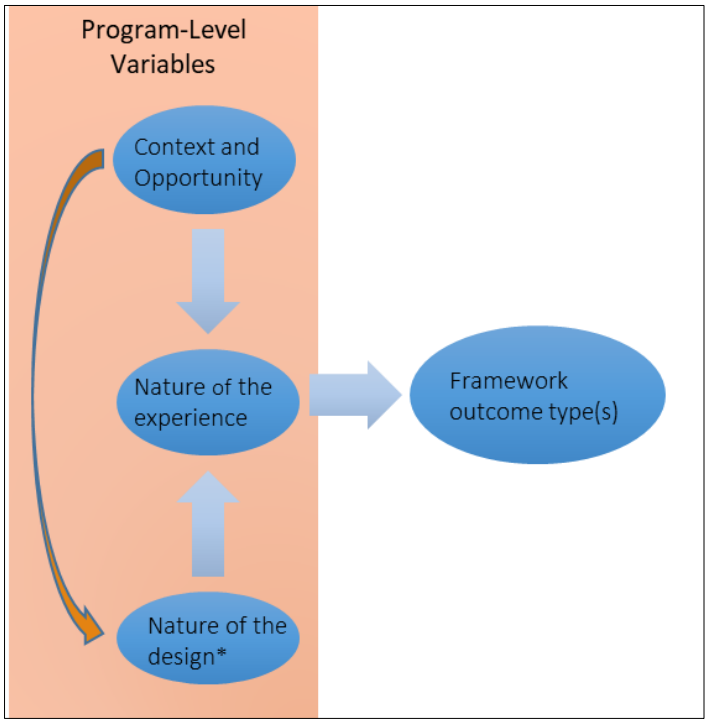


Figure 1. Framework outcome types result from program-level variables that lie outside the framework, requiring the framework to be nested within these variables.

These strategies include nesting the framework within a voices and choices planning process and clarifying underlying assumptions—primarily around the flexibility of category interconnectedness, outcome types within categories, and the meaning of their order.

Voices and Choices Planning for Self Determination: a context that insists on accountability for cultural differences, power, voice, and inclusiveness

Another way of making the framework more culturally responsive is to find ways of building into its structure co-production of program outcomes to include participant voice. In this case, outcomes would be selected and defined through mutual dialogue between program providers, participants, and other stakeholders whose voices will create programming most beneficial to participants and their communities. This effort would require attention to power such that no one voice overpowers another, with all perspectives considered, and decisions made with power differentials equalized. In this co-creation, participants—whose voice often is afforded the least power – bring their prior knowledge, interests, and values to program design.

Thus, the framework needs to be broad enough and flexible enough to accommodate (and not dictate) the voices and programming complexities they may seek. In other words, while providing structure for dialogue, the framework needs to allow room for processes and outcomes most meaningful to the community of learners it is serving—especially when those differ from preconceived notions of program designers.

Ultimately the goal of informal STEM education, or any educational experience meant to lead to durable learning and personal and social well-being, is to create an experience that generates or maintains intrinsic motivation, a sign of autonomous self-regulation and self-determination (Deci & Ryan, 2002). To account for how "context and opportunity" (as described above), influences self-determined program design and, in turn, the nature of the experience, we are recommending that the framework be used in the context of what we are calling Voices and Choices Planning for Self-Determination. This decision making process, or something similar, needs to be used at each stage of program planning (alignment with institution mission, program design, promotion, implementation, and evaluation).

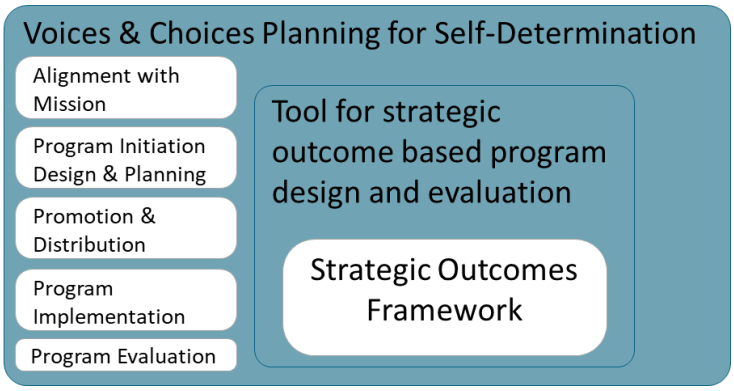


Figure 2. The framework is nested within Voices and Choices

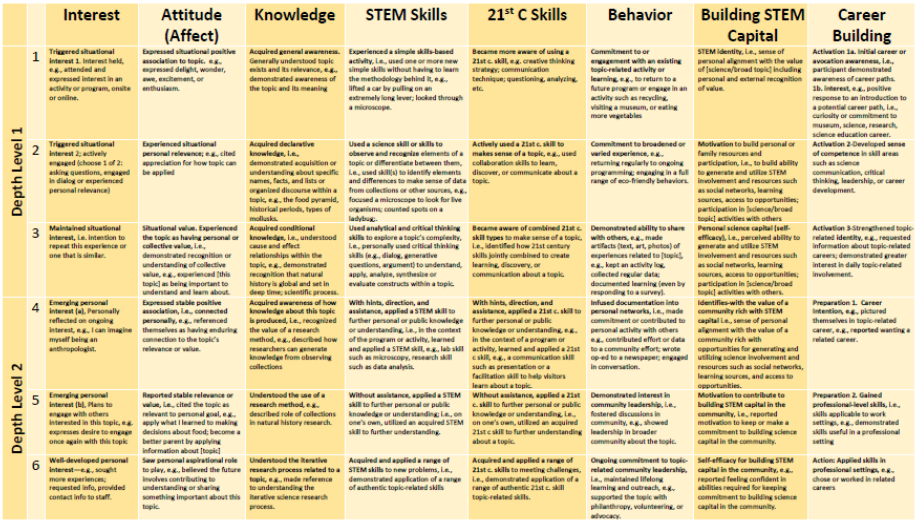
As envisioned, this process will involve two steps for each step of the program planning process (mission alignment through to program evaluation): (1) selection of voices that need to be present in decision making and (2) use of soft systems CATWOE methodology (Checkland & Poulter, 2006), which puts each voice alternately in the roles of customer, agent, and owner to reveal perspectives that encourage dialogue for making decisions about stakeholder inclusion and power sharing in decision making. Nested within voices and choices planning and the co-creative processes it engenders, the framework becomes a tool for program design and evaluation that is responsive and accessible to the wide diversity of learners they seek to serve.

This new framing will be tested in future iterations of the ISEE System, with the understanding that equity and justice work needs to be continual and something that always needs to be evaluated and addressed. The framework needs to function as a tool for internal reflection over time.

Framework Visual Reconfiguration

The conference revealed that the way the framework was presented reinforced some of the misinterpretations about its intent. In the original framework layout (Figure #3), learning categories appeared in columns. Within each column, rows represented a "progression" of six learning outcomes—with the top row representing an outcome that might be expected with the least amount of contact progressing step by step to the bottom row describing the most complex outcomes requiring the most amount and complexity of contact.

The original layout of the framework contributed to confusion about how the outcome statements were connected to one another. People misinterpreted the outcomes as being connected across each row (i.e., that Interest level one was related to Attitude level one, etc.). There was also a misunderstanding that the depth levels needed to be accomplished in order (i.e., that depth level two was a pre-requisite for depth level three).



		Interest	Attitude (Affect)	Knowledge	STEM Skills	21st C Skills	Behavior	Building STEM Capital	Career Building
Depth Level 1	1	Triggered situational interest 1. Interest held, e.g., attended and expressed interest in an activity or program, onsite or online.	Expressed situational positive association to topic. e.g., expressed delight, wonder, awe, excitement, or enthusiasm.	Acquired general awareness. Generally understood topic exists and its relevance, e.g., demonstrated awareness of the topic and its meaning	Experienced a simple skills-based activity, i.e., used one or more new simple skills without having to learn the methodology behind it, e.g., lifted a car by pulling on an extremely long lever; looked through a microscope.	Became more aware of using a 21st c. skill, e.g. creative thinking strategy; communication technique; questioning, analyzing, etc.	Commitment to or engagement with an existing topic-related activity or learning, e.g., to return to a future program or engage in an activity such as recycling, visiting a museum, or eating more vegetables	STEM identity, i.e., sense of personal alignment with the value of [science/broad topic] including personal and external recognition of value.	Activation 1a. Initial career or avocation awareness, i.e., participant demonstrated awareness of career paths. 1b. Interest, e.g., positive response to an introduction to a potential career path, i.e., curiosity or commitment to museum, science, research, science education career.
	2	Triggered situational interest 2; actively engaged (choose 1 of 2: asking questions, engaged in dialog or experienced personal relevance)	Experienced situational personal relevance; e.g., cited appreciation for how topic can be applied	Acquired declarative knowledge, i.e., demonstrated acquisition or understanding about specific names, facts, and lists or organized discourse within a topic, e.g., the food pyramid, historical periods, types of mollusks.	Used a science skill or skills to observe and recognize elements of a topic or differentiate between them, i.e., used skill(s) to identify elements and differences to make sense of data from collections or other sources, e.g., focused a microscope to look for live organisms; counted spots on a ladybug;.	Actively used a 21st c. skill to makes sense of a topic, e.g., used collaboration skills to learn, discover, or communicate about a topic.	Commitment to broadened or varied experience, e.g., returning regularly to ongoing programming; engaging in a full range of eco-friendly behaviors.	Motivation to build personal or family resources and participation, i.e., to build ability to generate and utilize STEM involvement and resources such as social networks, learning sources, access to opportunities; participation in [science/broad topic] activities with others	Activation 2-Developed sense of competence in skill areas such as science communication, critical thinking, leadership, or career development.
	3	Maintained situational interest, i.e. Intention to repeat this experience or one that is similar.	Situational value. Experienced the topic as having personal or collective value, i.e., demonstrated recognition or understanding of collective value, e.g., experienced [this topic] as being important to understand and learn about.	Acquired conditional knowledge, i.e., understood cause and effect relationships within the topic, e.g., demonstrated recognition that natural history is global and set in deep time; scientific process.	Used analytical and critical thinking skills to explore a topic's complexity, i.e., personally used critical thinking skills (e.g., dialog, generative questions, argument) to understand, apply, analyze, synthesize or evaluate constructs within a topic.	Became aware of combined 21st c. skill types to make sense of a topic, i.e., identified how 21st century skills jointly combined to create learning, discovery, or communication about a topic.	Demonstrated ability to share with others, e.g., made artifacts (text, art, photos) of experiences related to [topic], e.g., kept an activity log, collected regular data; documented learning (even by responding to a survey).	Personal science capital (self-efficacy), i.e., perceived ability to generate and utilize STEM involvement and resources such as social networks, learning sources, access to opportunities; participation in [science/broad topic] activities with others.	Activation 3-Strengthened topic-related identity, e.g., requested information about topic-related careers; demonstrated greater interest in daily topic-related involvement.
Depth Level 2	4	Emerging personal interest (a), Personally reflected on ongoing interest, e.g., I can imagine myself being an anthropologist.	Expressed stable positive association, i.e., connected personally, e.g., referenced themselves as having enduring connection to the topic's relevance or value.	Acquired awareness of how knowledge about this topic is produced, i.e., recognized the value of a research method, e.g., described how researchers can generate knowledge from observing collections	With hints, direction, and assistance, applied a STEM skill to further personal or public knowledge or understanding; i.e., on the content of the program or activity, learned and applied a STEM skill, e.g., lab skill such as microscopy, research skill such as data analysis.	With hints, direction, and assistance, applied a 21st c. skill to further personal or public knowledge or understanding, e.g., in the context of a program or activity, learned and applied a 21st c skill, e.g., a communication skill such as presentation or a facilitation skill to help visitors learn about a topic.	Infused documentation into personal networks, i.e., made commitment or contributed to personal activity with others e.g., contributed effort or data to a community effort; wrote op-ed to a newspaper; engaged in conversation.	Identifies with the value of a community rich with STEM capital i.e., sense of personal alignment with the value of a community rich with opportunities for generating and utilizing science involvement and resources such as social networks, learning sources, and access to opportunities.	Preparation 1. Career intention, e.g., pictured themselves in topic-related career; e.g., reported wanting a related career.
	5	Emerging personal interest (b), Plans to engage with others interested in this topic, e.g., expresses desire to engage once again with this topic	Reported stable relevance or value, i.e., cited the topic as relevant to personal goal, e.g., apply what I learned to making decisions about food; become a better parent by applying information about [topic]	Understood the use of a research method, e.g., described role of collections in natural history research.	Without assistance, applied a STEM skill to further personal or public knowledge or understanding; i.e., on one's own, utilized an acquired STEM skill to further understanding.	Without assistance, applied a 21st c. skill to further personal or public knowledge or understanding, i.e., on one's own, utilized an acquired 21st c skill to further understanding about a topic.	Demonstrated interest in community leadership, i.e., fostered discussions in community, e.g., showed leadership in broader community about the topic.	Motivation to contribute to building STEM capital in the community, i.e., reported motivation to keep or make a commitment to building science capital in the community.	Preparation 2. Gained professional-level skills, i.e., skills applicable to work settings, e.g., demonstrated skills useful in a professional setting
	6	Well-developed personal interest—e.g., sought more experiences; requested info, provided contact info to staff.	Saw personal aspirational role to play, e.g., believed the future involves contributing to understanding or sharing something important about this topic.	Understood the iterative research process related to a topic, e.g., made reference to understanding the iterative science research process.	Acquired and applied a range of STEM skills to new problems, i.e., demonstrated application of a range of authentic topic-related skills	Acquired and applied a range of 21st c. skills to meeting challenges, i.e., demonstrated application of a range of authentic 21st c. skill topic-related skills.	Ongoing commitment to topic-related community leadership, i.e., maintained lifelong learning and outreach, e.g., supported the topic with philanthropy, volunteering, or advocacy.	Self-efficacy for building STEM capital in the community, e.g., reported feeling confident in abilities required for keeping commitment to building science capital in the community.	Action: Applied skills in professional settings, e.g., chose or worked in related careers

Figure 3.: Original Framework Design

To clear up these points of confusion, we redesigned the framework visualization (Figure 4). Because people were inclined to read the table left to right, we reoriented the framework to have each row represent an outcome category; the columns now represent the depth levels. In addition, we gave each outcome category its own color to further distinguish them from one another. We also used color to visually signal the difference in depth levels by making the higher depth levels (i.e., Scope B) darker than the lower depth levels (i.e., Scope A). To clear up confusion around depth levels and pre-requisites, we increased the amount of white space around each outcome indicator statement to emphasize that they are individual statements and not a continuum. Further refinement of the graphic design (e.g., color pallete, font) is in progress.

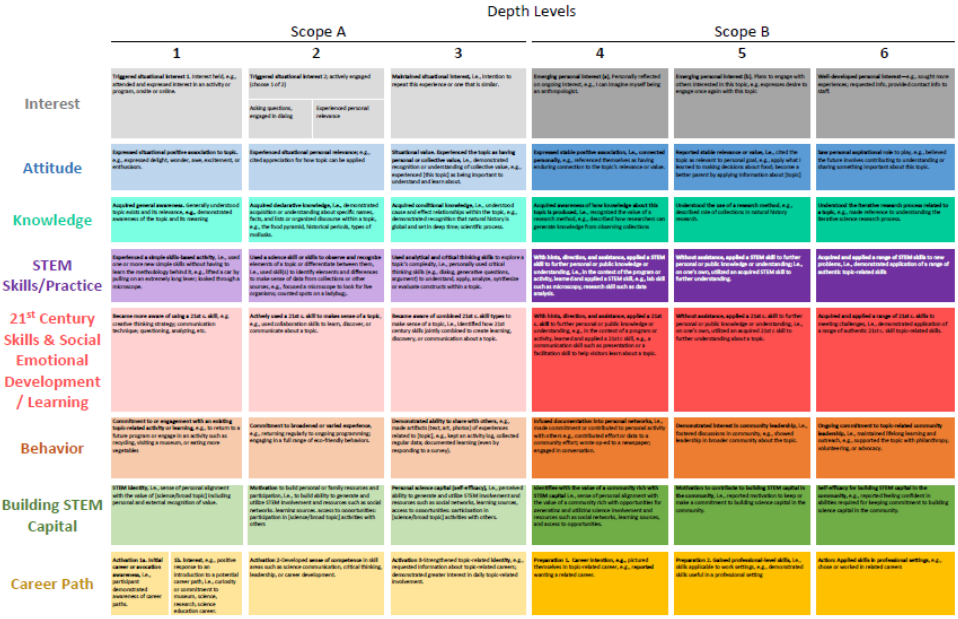


Figure 4: Redesigned Framework

Language to Describe the Framework

Categories

In response to conference participants' demand for outcome category definition and based on two accepted sources (Bell et al., 2009; Friedman, 2008), we have adopted the category definitions listed in Table 1.

Category	Definition
Interest	The state of wanting to be with, know or learn about something or someone.
Attitude	Personal reflections on concepts, processes, institutions, phenomena,
Knowledge -	Memory or understanding of facts, concepts, explanations, arguments, models
STEM Skills/Practice	Procedural aspects of knowing. Includes Manipulate, test, explore, predict, question, observe
21st Century Skills & SED/ L–	Self-concept and emotional knowledge; social awareness; emotion regulation; goal management; school work; relationship; decision making; critical thinking; creative thinking; et al.
Behavior	Demonstrations of, assessment of, change in, or exercise of activity related to a STEM topic
Building STEM Capital	What science you know, who you know, how you think, what you do; what may help transform positive attitudes and interest in STEM into actions that make future science participation more likely
Career Building	Used as a measure of long-term impact of STEM education and engagement and as a measure for scientific identity development

Table 1. Framework category definitions.

Outcome Types and their arrangement within Categories

Within each category, outcome statements describe general buckets and are not prescriptive in and of themselves. Thus, we have begun referring to them as outcome "types." Outcome "types" need to be generalizable enough for institutions to define outcomes within those types in ways that are in line with the institution's mission and philosophy.

The order of outcome types function, not as a pathway, but as a way of describing the intent of programming. Although ordinal, outcome types function independently. One outcome type is not dependent on achieving the type that occurs before it. On the other hand, in some cases, not all, programming may want to encourage participants to see themselves as moving from one place in the order to another.

Whatever the outcome types, each works in conjunction (supporting vs. primary) with others both within the same category and across categories.

The theoretical and empirical rationale (theory and research) behind the order of outcome types needs to be readily available with the framework. d

Programs intending outcome types at the beginning, simpler depth levels tend to be shorter, lower contact experiences that reach larger numbers of people. Those at more complex depth levels happen with more contact repeated over time and reach smaller, more selected numbers.

Not all categories have the same number of outcome types between the anchors that define the first and last of outcome types in the order.

Outcome-Type Specification to Create an Outcome Statement

To generate useful outcome statements, practitioners need to specify outcome types to reflect a program's topic, intention, and learners. Specification can happen both in terms of the topic being addressed (e.g., "scientific process," "frogs," "climate change," "STEM") and sub-category (e.g., "wonder" for interest type, "critical thinking" for 21st century skill type, or "recognition" for knowledge type). In and of itself, outcome type specification generates important planning, implementation, and evaluation deliberation.

"Depth" and "Complexity" as an explanation of outcome order

Words to describe the order included "level" "depth" "impact array" and "graduation." As the conference progressed, the idea of greater and less "complexity" emerged as a way of understanding the ordinal arrangement of outcome types within categories: from simpler to more complex. If "complexity" is used to describe the order, it will be important to clarify complexity and to what the complexity refers. In other words, does the order reflect programming complexity, outcome complexity, or both?

Interconnection of Outcome Categories and how the framework relates to logic, theory of change, and pathway models.

Adopted in the context of the conference was the use of macramé as a metaphor to portray educational programming and the outcomes as a fabric of knots made from strands of learning categories tied together purposefully at varied depths (or outcome types) to create a learning design. This concept of knots tied from various outcome strands helped participants consider the framework as offering structures for purposefully considering how various outcome intentions work together to reflect or intentionally create multi-dimensional programming.

These "knots" can reflect logic models, more complex theory of change models, or pathway models. Logic models delineate resources, activities, and outputs that lead to an outcome, which can in turn lead to a longer-term outcome (Allen et al., 2008). Theories of change (Mayne, 2015) and their corresponding pathway models (Urban, Hargraves, et al., 2021) explicate series of various types of outcomes that contribute to other overarching outcomes, each with its own logical explanation of the conditions necessary for achieving the outcome. Informal STEM education evaluation typically utilizes logic modeling (Dierking, 2008). This framework supports one or the other, sometimes both (Meyer et al., 2021). With guidance from this Strategic Outcomes Framework, practitioners can construct these models from the eight category "strands," with knots tied at various outcome type depth levels within each strand. In other words, each category provides choices for outcome types at varying depth levels. These outcome types can be selected as primary or supporting outcomes, and within those, as intended or reach outcomes. Once selected they need to be specified to reflect stakeholder voice and choice in relation to the learning.

Primary, Supporting, and Reach Outcomes.

Outcome selection and consequent program planning can be conducted with more clarity after dividing outcome intention into three areas: primary, supporting, and reach. Primary and supporting occur first prior to selecting the reach outcome(s). The primary outcome defines the program or exhibit's ultimate educational purpose. The supporting outcomes describe learning that will lead to the primary outcome. Both primary and supporting outcomes describe intentions for all participating learners. In contrast, reach

outcomes describe potential (e.g. “expect to see,” “like to see,” and “love to see” as used in outcome mapping; Earl et al., 2001). Primary and supporting outcome selection can significantly alter program design and lead to rich discussion. Consider, for instance, the difference in programming (and the discussion that determines it) when the primary outcome of a “climate change awareness” program is interest, attitude, knowledge, skill, or behavior. And then consider the difference in programming when choosing which of these categories support the selected primary outcome. How does the programming further alter if one of these categories is defined as a “reach” outcome?

Selecting Categories: Framework relationship to theory of change, program logic, and context.

The framework can both help build and be supported by theory of change, program logic, and context.

Theory of change and logic models can work in tandem, especially in the context of assessing systemic influences with communities of color (Meyer et al., 2021). Theories of change outline the series of outcomes that describe a causal pathway toward an ultimate intended outcome (Mayne, 2015). As such, the framework can be used to help articulate supporting and primary outcomes. In turn, logic models describe theory of action that leads to an outcome. These theories outline resources, activities and outputs that lead to an outcome (some logic models include a series of short-, mid-, and long-term outcomes that describe a theory of change). Ordinal organization of outcomes within framework categories contribute to logic model construction by contrasting outcomes according to the complexity of programming required to produce them. Thus, practitioners using the framework can contrast an intended outcome type with other outcome types within a given category to better match the intended outcome to programming activities and contact with those activities.

Clarification of framework use

As illustrated by its placement within the ISEE system, whether within or outside of the context of the ISEE system, the framework is meant to be useful throughout the program design-promotion-implementation-evaluation process. As such, and in conjunction with a method for optimally involving relevant voices in decisions it requires (e.g., voices and choices planning in the ISEE system), the framework leads to asking pertinent questions and guidance involved with the design of:

- institutional education goals and portfolio of programs that serve to meet those goals;
- program-specific goals and realistic objectives;
- ways to promote a program in relation to the outcomes it is meant to produce;
- ways to implement program activities in ways that meet those objectives; and
- ways to evaluate learning within the program.

Note that the framework generates discussion and guidance of educational goals and learning outcomes. Broader outcomes such as affecting public policy or general community awareness are beyond the scope of this framework.

The framework is meant to be used within institutions, not necessarily across them. The calculation is one of specificity over generalization and dilution. We think it is more important at this point to allow institutions and their value holders to customize the framework to their individual needs and approaches than to create standards that are comparable – but potentially less meaningful and more difficult to apply – across all institutions.

Further Questions/Future Work

The conference also served to layout an agenda for readying the framework for dissemination. Agenda topics include further clarification and definition of the framework categories; user support, accounting for outcomes not in the framework, and future research.

Further clarification and definition of STEM Categories

Categories and outcome types within them

For each of these categories and outcome types within them research documentation needs to continue and additional research on realistic, contact-specific expectations within categories needs to occur. The outcome types need to be interrogated for ungrounded assumptions. Category definitions need to be challenged and broadened from the perspective of diverse cultural responsivity, equity, and inclusivity. Moreover outcome types must be written in ways that preclude a deficit-based approach.

Notes for each of the categories are listed below.

Interest

Because of the great deal of research conducted on interest levels and how they differ and develop from situational to personal (Renninger & Hidi, 2011), the interest category provides a model for how the other categories can be researched and depth levels within them defined. Future work needs to consider how strategies for supporting interest differ for each depth level.

Attitude

Once specified, attitude outcomes can cover a wide range of informal STEM programming. For example, if "having fun" is included in the definition of attitude, the category becomes highly useful for programming seeking to generally affect attitude toward STEM involvement.

Depth levels within attitude need review in light of expectancy-value theory and how they can be specified in ways that include the associated perceived value of the attitude being affected. In this way, attitude levels will need to include co-definition with participants.

Specified outcomes and their corresponding indicator statements will also need to reflect both value and relevance as perceived by evaluands. These statements need to consider how participants describe concepts, processes, institutions, phenomena, etc. that hold these value and relevance qualities.

Knowledge

Knowledge outcome types, their definition, and especially their specification needs to allow for the range between bodies of knowledge and knowledge pieces. Moreover specification options need to be broad enough to encompass relationship to knowledge across cultures, venues, values, etc. Guidance for outcome specification will be particularly important to allow users to discriminate between knowledge.

STEM Skills (and/or STEM Practices)

Discussion has led us to the possible necessity of expanding STEM skill to include STEM practice. For example, STEM practice may be particularly useful as an outcome that supports another outcome category. Such was the case in one participant's description if a 6th grade class studying statistics where youth experienced the practice of merging their personal and community knowledge with STEM knowledge to build new and deeper understandings of statistics:

"…The teacher purposefully used the context of evaluating playground safety to teach about environmental statistics. He hoped that this would "create spaces" for students to leverage on their knowledge of how playgrounds work, and the issues that matter in playgrounds. Students used their knowledge to not only deepen their understanding of statistics but also to shift the collective outcomes for the whole class, including, arguing for a "real" evaluation that makes a difference, changing evaluation criteria to match more realistic concerns…."

In this case, STEM practice was a supporting outcome in service to one of a number of primary outcomes: knowledge of statistics, attitude toward STEM utility, attitude toward STEM practice, and possibly others. In this way, considering STEM practice as a supporting outcome invites program strategy that encourages use of everyday knowledge to build deeper understanding of STEM topics, issues, and processes. Awareness of STEM practices becomes knowledge and skill to further learning. For example, naming practice such as "I am doing STEM" and attitude "I have fun doing STEM" or "I recognize my daily knowledge and practice of caring for sled dogs involves STEM knowledge and STEM Skill" leads to "I can apply my knowledge attitude and skill to learning more."

With the adoption of the concept of "practice" to augment skill, we will need to provide guidance for identifying both "practice" and "skill" in a single program. Definition needs to clarify the difference and their relation to each other.

Moving to include "practice" as an outcome has important advantages. On the other hand, some programs intend to teach specific STEM skills. In some cases, practice leads to skill. In other cases, practice and recognition of practice is in and of itself the intention. It will be important to review the "skill" outcome types for their relevance to "practice." The outcome types potentially still work.

SED/L & 21ST century skill

The 21st century skill category has been merged with social-emotional development and learning (Fenichel & Schweingruber, 2010). This merger will require scrutiny for applicability of outcome types.

Following STEM skill, the current depth levels for this category start move from skill acquisition to application. One way of determining the utility of these outcome types is to map them, along with supporting outcomes from other categories, onto processes for socio-scientific reasoning (Zeidler, 2016). This highly necessary exercise will provide a rigorous test of the framework's flexibility. It will also reveal important guidance for selecting outcome types and specifying outcomes.

Behavior

Questions remaining to answer include:

- Do the outcome types work for all sets of behaviors?
- What outcomes generate casuistic reflection about actions/ behaviors related to complex issue education?
- What strategies supporting transfer of behaviors beyond immediate within-program outcomes?

STEM Capital

We will review STEM Capital outcome types to be more in line with research on STEM Capital, as suggested in the table below.

Table 2. Suggested revisions for STEM Capital outcome types.

Original Depth Levels	Breadth	Access & Acquisition	Contribution
	STEM identity	1. STEM Topic Identity	4. Identifies with the value of a community rich with STEM capital
	Motivation to build STEM Capital	2. Motivation to build STEM capital personally or within family	5. Motivation to contribute to building STEM capital in the community
	Personal STEM capital (self-efficacy	3. Personal STEM Capital (self-efficacy)	6. Self-efficacy for building STEM capital in the community
Revised Depth Levels		Personal Development	Focus on equity/power
	STEM Capital	1. Building personal STEM capital (beyond knowledge and dispositions)	4. Recognizing & valuing the community as rich with STEM Capital – broadening what counts as STEM
	STEM Identity	2. Developing/supporting STEM identity	5. Developing/supporting more inclusive STEM identity—broadening who counts as STEM
	STEM agency	3. Developing/supporting personal STEM self-efficacy	6. Utilizing STEM Capital to take action to improve community

These revisions depend on a definition of STEM Capital as the maximized equitable conception of STEM as defined by transformative power, representation of underserved interests, asset-based approach, collective orientation; centrality of participation as key, redistribution of resources, and long-term time frame (Figure 5).

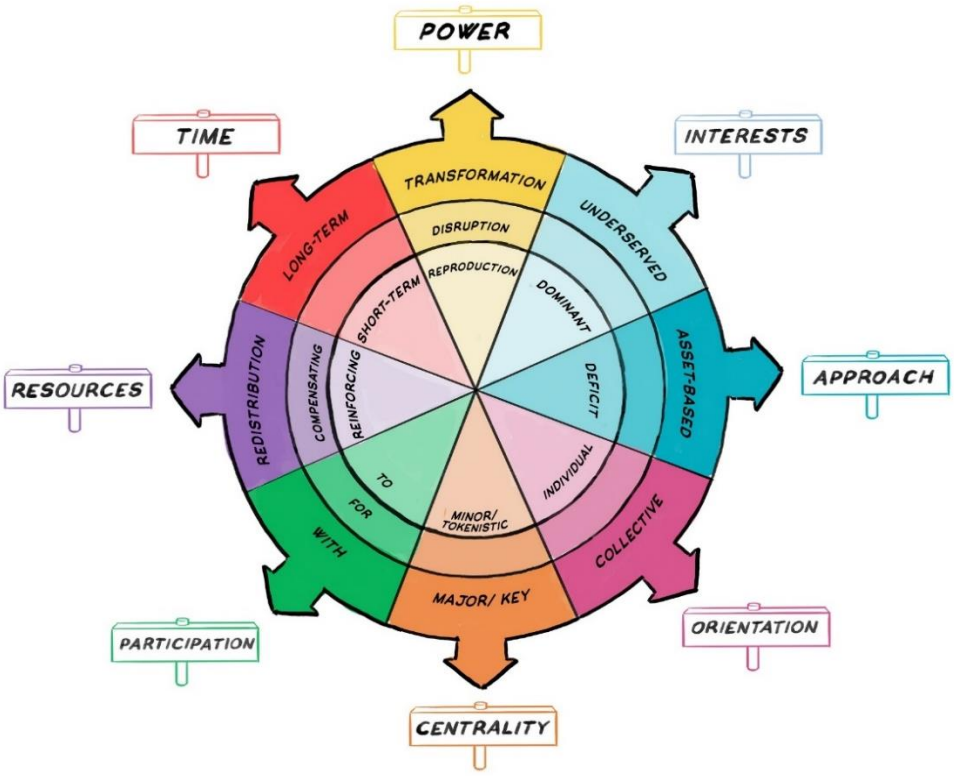


Figure 5. Equity compass for defining STEM Capital

Career Path

As part of his conference participation, career path content expert, Robert Tai, recommended changes to Career Path outcome types (bold type in Table 3 below). We will adopt those changes into the framework.

Table 3. Career Path suggested outcome type revisions

Breadth		Interest	Activation	Application
Depth levels	Activation	1a. Career or avocation **awareness** 1b. Career or avocation **interest**	2. Developed sense of competence and seek to further informal learning	3. Strengthened topic-related identity and engaging in informal teaching activities
	Preparation	4. Career intention in educational planning	5. Gained professional level skills and educational trajectory	
	Outcome			6. **Applied skills** in a professional capacity

Additional questions to consider include:

- What differences in the proposed progression might exist for non-White, non-Male, non-Cisgender individuals?
- How is anti-science thinking impacting youth engagement?

Centering learners vs. Centering programs: an important conceptual issue to resolve.

Are we centering the learners, or the programs when thinking about outcomes? (New Ideas to Consider board) [on the New Ideas board, someone responded, "yes."] This issue emerges when identifying depth level.

User Support

Many comments, recommendations, and discussions included both general and specific suggestions for a manual that would support the framework's use. In addition, processing these comments led to even more ideas/visions for the manual. For that reason, we have begun creating a user manual outline with notes that come from all of these sources (Appendix D).

Guidance will be required to help users understand these relationships. User training needs to include guidelines for and examples of how programs will look different based on different configurations of outcomes. Definitions with examples and the ability to designate primary and supporting categories will help. Delineation of outcome types is not enough; educators need to make them explicit and will require guidance for specifying discrete outcomes within outcome types. Such guidance will make clear that outcomes in the framework rarely operate alone. In most cases, a combination of outcomes work together to achieve an intended learning outcome or combination of outcomes. These combinations or "clusters" might prove to be useful as program types. Outcome types can function in both primary and supporting roles. Some categories necessitate outcomes in other categories, e.g. sharing knowledge necessitates presentation skills.

Additional outcome-specific notes that should be included in a user manual include the following:

- STEM identity can be related to STEM capital.
- STEM skills and 21st century skills can be intertwined.
- Attitude toward STEM is connected with STEM career choice, recognition of STEM career value, self-efficacy, STEM identity, and sense of belonging.
- STEM careers can involve critical thinking. Either can be primary with one supporting the other.
- In addition to an affective component, attitudes have a behavior component. E.g., attitude toward snakes might be supported by skill (holding a snake properly) and behavior (regularly approaching snakes properly).
- Critical thinking can support each of these categories and, for each, would be defined differently.
- Interest can support all the categories.
- Self-efficacy can support each of the other categories.
- Learning can occur without interest.
- STEM capital is related to attitude (value).

Definitions of outcome types and depth levels will need to reference extant literature. They also need to leave room for co-definition with participants, and the "perception" of value needs to be made clear.

Outcome Selection and Specification

Across all categories, users will need examples of specification and perhaps even a menu of possible terms that might be useful.

Accounting for Outcomes Not in the Framework

Each of the following concepts needs to be evaluated for how they can be operationalized within the current framework structure, or if they need to alter the framework structure. The user manual will include guidance for the types of questions that need to be answered to locate the concepts/topics below into the framework. For example, with programs addressing social justice issues and systems change, at what level does the program function? Is it awareness (knowledge) of an issue? Skills to address the issue? Behavior change? The manual needs to provide examples of relevant programs, how they operationalized their intended outcomes, and the questions they needed to answer to be able to arrive at outcome type selection and specification. These topics include:

- Social justice issues and system change
- Social-emotional development (can it be part of 21st century skills or does it need its own category?)
- Self-efficacy and sense of competence
- STEM identity
- Building STEM community; sense of belonging
- STEM literacy
- Beliefs and values
- Empathy
- Constructivist, serendipitous learning experiences
- Applying STEM to address social and other non-STEM issues
- Transformation as an outcome type (in addition to acquisition and assimilation

Use for Evaluation Purposes

Perhaps the most important issues to be addressed to prepare the framework to be useful for evaluation purposes, is the operationalizing of each outcome type. The process will, among other strategies, involve adapting existing shared measures to the framework outcome types.

A serious limitation of logic model construction or theory of change models is that they assume an "average" starting point for all participants. Despite widespread adoption of logic model planning and evaluation in informal learning, this limitation is particularly problematic in informal learning where so many activities invite learners across a full range of entrance abilities. One way of addressing this problem could involve goal attainment scaling, an evaluation method often used for wrap-around services where programming differs based on the needs of the individual participant (Marson et al., 2009). In this case individual learners might use the framework to select the outcomes they would personally like to achieve. Programming would then be evaluated based on progress toward selected outcomes. There has been some discussion at COSI about using the framework to assist with goal attainment scaling so that educators can use the framework as a guide for helping individual participants set their own learning goals.

This is a challenge that needs to be addressed in future framework use development.

Future research

Following is a bucket list of issues still to be addressed in the framework's future development research:

Unit of analysis: individual, family, community

Explanation needs to be provided for how to use the framework categories and outcome types at varied units of analysis individual, family, and community levels.

Outcomes within learning ecosystems

Another aspect of learning outcomes to be considered is how informal educational learning fits as part of a whole learning system, including both formal and at-home settings. How do these outcomes interface with the wider web of the learners' own outcomes or those of other adults or peers in the learners' lives (e.g., teachers, parents, caregivers, etc.)?

Need for buy-in/awareness from funders and institution leadership

Need for buy-in and communication with both leadership and funders also emerged as an important element to framework utility. Some of the issues included:

- Dealing with pressures for producing outcomes not included in or not easily defined by the framework.
- Making the framework relevant to and connected to institutional leadership concerns.
- Assuring that funders, development officers, CEOs, and other stakeholders don't mistake the relative value of each column or row.

Staff capacity

As a corollary to the framework's contribution to building staff evaluation and planning capacity, there is a concern that its complexity in both its form and use may make it less usable by staff.

Framework Flexibility

The framework needs to accommodate programming across:

- All age groups and differing needs within groups
- Venues (in-school, out of school, in-museum, out-of-museum)
- Digital/virtual programming
- Multiple instructors
- Simple and more complex needs (i.e., provide a simple entrée for less-experienced practitioners and more details for those who need them)
- Ability to locate "where I was before and where I am now" on a given depth level or, where and when appropriate, between levels

References

Allen, S., Campbell, P. B., Dierking, L. D., Flagg, B. N., Alan, J. F., Garibay, C., Korn, R., Silverstein, G., & Ucko, D. A. (2008). *Framework for evaluating impacts of informal science education projects: report from a national science foundation workshop* (A. J. Friedman (Ed.)).

Allen, S., & Peterman, K. (2019). Evaluating Informal STEM Education: Issues and Challenges in Context. *New Directions for Evaluation*, *2019*(161), 17–33. https://doi.org/10.1002/EV.20354

Aronowitz, S. (2004). Against Schooling. *Social Text*, *22*(2), 13–35. https://doi.org/10.1215/01642472-22-2_79-13

Bell, P., Lewenstein, B., Shouse, A. W., & Feder, M. A. (2009). *Learning science in informal environments: People, places, and pursuits* (P. Bell, B. Lewenstein, A. W. Shouse, & M. A. Feder (Eds.)). The National Academies Press. http://books.google.com/books?hl=en&lr=&id=vdOIlY7WdJUC&oi=fnd&pg=PT1&dq=informal+education+exhibit+influence+on+behavior&ots=PQ_InA5U8t&sig=mj1aviIrKNkc1cIRd4g8TCdgRG0

Checkland, P., & Poulter, J. (2006). Learning For Action: A Short Definitive Account of Soft Learning For Action. *John Wiley Son Ltd West Sussex England*, xxiii, 200 p.

Deci, E. L., & Ryan, R. M. (2002). *Handbook of Self-Determination Research: Theoretical and Applied Issues*. University of Rochester Press.

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, *26*(3–4), 325.

Dierking, L. D. (2008). Evidence and categories of ISE impacts. In J. Friedman, Alan (Ed.), *Framework for evaluating impacts of informal science education projects*. National Science Foundation Directorate for Education and Human Resources DRL.

Earl, S., Carden, F., & Smutylo, T. (2001). *Outcome Mapping: Building learning and reflection into development programs | Better Evaluation*. International Development Research Centre. https://www.betterevaluation.org/resource/outcome_mapping

Fenichel, M., & Schweingruber, H. A. (2010). *(2010). Surrounded by science: Learning science in informal environments*. National Academies Press.

Friedman, A. J. (Ed.). (2008). *Framework for Evaluating Impacts of Informal Science Education Projects | InformalScience.org*. https://www.informalscience.org/framework-evaluating-impacts-informal-science-education-projects

Garibay, C., & Teasdale, R. M. (2019). Equity and Evaluation in Informal STEM Education. *New Directions for Evaluation*, *2019*(161), 87–106. https://doi.org/10.1002/ev.20352

Innovation Network. (2012). *Point K Learning Center*.

Marson, S. M., Guo Wei, & Wasserman, D. L. (2009). A Reliability Analysis of Goal Attainment Scaling (GAS) Weights. In *American Journal of Evaluation* (Vol. 30, Issue 2, pp. 203–216). https://doi.org/10.1177/1098214009334676

Mayne, J. (2015). Useful Theory of Change Models. *Canadian Journal of Program Evaluation*, *30*(2). https://journalhosting.ucalgary.ca/index.php/cjpe/article/view/31062

McLean, K., & Pollock, W. (2011). *The Convivial Museum*. ASTC-Association of Science-Technology Centers.

Meyer, M. L., Louder, C. N., & Nicolas, G. (2021). Creating With, Not For People: Theory of Change and Logic Models for Culturally Responsive Community-Based Intervention. *Https://Doi.Org/10.1177/10982140211016059*. https://doi.org/10.1177/10982140211016059

Renninger, K. A., & Hidi, S. (2011). Revisiting the Conceptualization, Measurement, and Generation of Interest. *Educational Psychologist*, *46*(3), 168–184. https://doi.org/10.1080/00461520.2011.587723

Sneider, C., & Noam, G. G. (2019). Sneider, C., & Noam, G. G. (2019). The common instrument suite: a means for assessing student attitudes in STEM classrooms and out-of-school environments.

Connected Science Learning, 11. *Connected Science Learning*, *11*.
Teasdale, R. M. (2021). Evaluative Criteria: An Integrated Model of Domains and Sources: *Https://Doi.Org/10.1177/1098214020955226*, *42*(3), 354–376. https://doi.org/10.1177/1098214020955226
The Netway. (2021). *The Netway Program & Evaluation Planning*. https://www.evaluationnetway.com/
Urban, J. B., Archibald, T., Hargraves, M., Buckley, J., Hebbard, C., Linver, M. R., & Trochim, W. M. (2021). Introduction to Relational Systems Evaluation. *New Directions for Evaluation*, *2021*(169), 11–18. https://doi.org/10.1002/EV.20444
Urban, J. B., Hargraves, M., Buckley, J., Archibald, T., Hebbard, C., & Trochim, W. M. (2021). The Systems Evaluation Protocol for evaluation planning. *New Directions for Evaluation*, *2021*(169), 31–45. https://doi.org/10.1002/EV.20443
Wasserman, D. L. (2010). Using a systems orientation and foundational theory to enhance theory-driven human service program evaluations. *Evaluation and Program Planning*, *33*, 67–80. https://doi.org/10.1016/j.evalprogplan.2009.06.005
Zeidler, D. L. (2016). STEM education: A deficit framework for the twenty first century? A sociocultural socioscientific response. *Cultural Studies of Science Education*, *11*(1), 11–26. https://doi.org/10.1007/s11422-014-9578-z

Appendix A. Recommendations from Research Consultant Martin Storksdieck, with responses referenced in this white paper

In addition to the eight content expert research consultants, we also invited Martin Storksdieck to attend the conference as an informal education researcher who could provide feedback and recommendations for the framework's future development to ensure it is useful to informal education practitioners. Following is his verbatim response. This white paper supports his observations with analysis of participant comments and feedback. It also addresses his recommendations.

What supports do we need to create so that evaluators and educators can best use it?
Before I can answer this question, I need to acknowledge that feedback on the Framework may require some rethinking about design and purpose. The concept of a progression was questioned by experts across domains, as were the domains themselves:

(1) The term "learning progression" already exists and has a specific connotation that deviates from how the term was used in the Framework. This is partially a semantic, and somewhat a conceptual issue. Renaming the degrees to which a person "gets better" or is more capable within a domain can solve the problem to a degree. The other issue is that in some domains (particularly affective and dispositional ones), the concept of a linear improvement ("progression") was questioned.
(2) Reviewers felt that some domains (the columns in the Framework) needed rethinking within (what do they mean? What would be included?) and across – suggestions ranged from lumping skills to better positioning concepts such as "STEM capital".
(3) Individuals start at different points, come to the experience with their individual portfolio in which "depth" differs between outcome categories. The Framework's biggest problem is its need to define a statistically archetypical participant, and then needing to deal with diversity within categories. That's a tough problem. If the purpose of a short-term engagement can only be to trigger interest, what do you do with those who are already interested? You now have to distinguish triggering from sustaining (or such), and assign it to the exact same experience, depending on who the person is. Which makes the Framework 3-dimensional.

The other issue was purpose and target group: Is the Framework a precise guidebook for evaluation and measurement, one that allows to define exactly what to measure due to the nature of the experience that can be mapped against the Framework? Or is the Framework a planning instrument, one that allows experience designers in thinking what is possible, and to help guide decisions about goals and objectives in realistic and accomplishable ways? The meeting seemed to have pointed to the latter, given the many complexities and perceived needs for changes in the Framework. The level for needed "validity" of the Framework is much lower if it serves as a planning and design heuristic, and primarily serves to guide thinking, than if it serves as a tool to shape evaluation and research studies.

From here on I will assume that the Framework is a two-dimensional planning tool that allows to map an experience design against various outcome domains, whereby the domains distinguish depth of these outcomes. In short, it not only allows to ask what outcomes are possible or desirable from a particular experience design, but taking time and/or depth of engagement into account, also speaks to how much within each outcome category can be accomplished on general principle. Note, though, that at least for now, it would be important to not insinuate that all "rows" in the Framework indicate corresponding levels of depth across outcome domains. At least for now reviewing experts did not find sufficient

evidence that there was equivalency, and the team acknowledged as much. However, the table format of columns and associated rows sends this message very strongly and must therefore be rethought.

So back to the supports:

Where should the framework go next?
I think the conference gave an abundance of guidance on how to make adjustments and improvements, though it wasn't always clear how to make these changes. The use cases as presented were somewhat misleading since they could not take the critique of experts into account. In fact, the use cases show the danger of releasing a Framework to practitioners: it will likely get used "as is" (independent of any need for change) because it fills a void or gap. Learning Styles and Multiple Intelligences are just two examples of memes running rampant in practitioner communities, and they are just the tip of the iceberg. Alas, sometimes imperfect guidance can be better than none. So here is where I can see the Framework go:

A) Respond to constructive critique and create at least two alternates to the current version, playing with different forms of representation.
B) Send back to original group for feedback. Keep in mind that some experts questioned the Framework at a more fundamental level and might never be satisfied.
C) Couch initially as planning tool. It is a lower bar… For a planning tool it needs to foreground the notion of mapping an experience against the Framework to understand what can reasonably be accomplished and claimed, or to mark in the Framework what is desired and then choose the corresponding experience. Make it clear that the reality of experience design is an iteration between creative "let's do this" and thoughtful "but why and for what purpose". The Framework, used in this way, would then need an interactive website with support for implementation, training and training manual (if one can be created), and some form for exchange by practitioners who used it. The www.islframework.org is an example of this that was NOT sufficiently funded to get everything done.
D) In order to make the Framework a tool for research and evaluation, the definitions of each outcome domain needs sharpening, and much more thought needs to go into creating the "depth bins" (the stages of getting or being better at it). I think you need to go back and ask yourself a different "row" question that is practitioner oriented: what if I have little time (the typical museum 30 sec to 2 min), all the way to a summer program.
E) If it is meant to be used as a research/evaluation tool, then maybe a focus on the depth is less needed – it is just too controversial. However, it would be useful to link outcome categories to constructs that represent those and could be measures, and measures themselves (but note that others already do that, so maybe link to those efforts). The Framework would then become a visual presentation of a outcome measure repository.

What people, organizations need to learn about it in order to most effectively disseminate it?
The people you invited are a good start. A designated website. The typical dissemination efforts. It is hard to answer without knowing who you are already targeting at ASTC, ACM, AZA, AAM, VSA, AEA, etc. NARST, AERA, but more AEA and VSA would be a fora for further critical discussions. ASTC sessions and workshops would be good places to explore usefulness with practitioners.

What research needs to be conducted to best support it?
That is not clear to me right now since it depends on your next moves. If you go the route of a planning and design tool, then research is about usability and usefulness and the target are practitioners. If you want to refine it and turn it more into a theory and conceptual framework, then your biggest research

issue is to define “depth” or what it means to get better in those domains, building off of the work the experts provided. Maybe select a few and work with those experts.

Appendix B. Conference Participants

ISE PROFESSIONALS

Educators, evaluators and researchers in informal STEM education who might use the framework

Andrew L. Aichele, COSI

Sarah Dunifon, Improved Insights LLC

Marlena Jones, CASE/Carnegie Science and DC STEM Network

Kathayoon Khalil, Zoo Advisors LLC

Judith Koke, Institute for Learning Innovation

Karen Lee, National Park Foundation

Priya Mohabir, New York Hall of Science

Giuseppe (Pino) Monaco, Smithsonian Institution

Fanuel Muindi, STEM Advocacy Institute and Harvard University

Karen Peterman, Karen Peterman Consulting Company

Allison Price, Lincoln Park Zoo

Jennifer Rehkamp, Association of Children's Museums

Nelda Reyes, AB Cultural Drivers

Edna Tan, University of North Carolina, Greensboro

Carrie Tzou, University of Washington, Bothell

Mauricia A. Vasquez, Frost Museum of Science

Latasha Wright, BioBus, Inc.

OUTCOME RESEARCHERS

Experts in the developmental theory that underlies the outcomes

Cameron D. Denson, Mwenda Kudumu, (Attitude), North Carolina State University

Spela Godec (STEM Capital), University College London

Joe Heimlich (Behavior), COSI Center for Research and Evaluation

K. Ann Renninger (Interest), Swarthmore College

Tiffany-Rose Sikorski (Knowledge), George Washington University

Robert Tai (Career), University of Virginia

Dana Zeidler (21st Century Skills),

University of South Florida

Angela Calabrese Barton (STEM Skills), Michigan State University

ISE RESEARCH SYNTHESIZER

Martin Storksdieck, Oregon State University

CONFERENCE STEERING COMMITTEE

Patricia J. Allen, Institute for the Study of Resilience in Youth (ISRY)

Jennifer Anstadt, National Museum of Natural History

Karyl Askew, Karyl Askew Consulting, LLC

Kristin Lewis-Warner, PEAR
Gil Noam, Institute for the Study of Resilience in Youth (ISRY)
Priya Mohabir, New York Hall of Science
Colleen Popson, National Museum of Natural History
Cathlyn (Cat) Stylinski, University of Maryland
Jessica Takacs, COSI Center for Research and Evaluation
Deborah Wasserman, COSI Center for Research and Evaluation

Appendix C. Evaluation Report

Strategic Outcome Progressions Conference: Exploring a Framework for Measuring Informal Education Outcomes and Institutional Impact (NSF AISL #2039209)

Evaluation Section – What Happened?

Cathlyn Davis Stylinski – University of Maryland Center for Environmental Science
Karyl Askew – Karyl Askew Consulting, LLC

Introduction

"What happened" during the conference was captured through evaluation co-led by two external consultants, Cathlyn Davis Stylinski (University of Maryland Center for Environmental Science) and Karyl Askew (Karyl Askew Consulting, LLC). This Evaluation Team used a collaborative and equitable approach for this evaluation. As such, they focused on two sets of evaluative questions:

- *How successful was the team at diversifying participation and input from multiple perspectives? Did participants feel their voice and ideas were heard and valued throughout the process?*
- *How did the conference structure and process affect how participants felt about the Framework?*

Additionally, the Evaluation Team provided feedback throughout the development and implementation process through a lens of inclusion. Finally, they took care to protect the confidentiality of the participants by not releasing any individualized responses; instead the evaluators provided their reflections based on participant-provided data along with their direct observations captured via field notes. Data collection and findings are described in the following sections.

Data collection

Both two evaluators participated regularly in Steering Committee meetings, in which the Leadership Team (who led facilitation of the conference) along with the CREED members and other advisors (who participated and supported conference activities) worked on preparations for the spring 2020 conference. During this planning process, the Evaluation Team encouraged and supported the Leadership Team in the articulation of criteria that explicitly attended to a recruitment pool of ethnically, geographically, and organizationally diverse conference participants. Both evaluators also attended all three synchronous conference sessions (April 8, 15 and 26). At all of these meetings, they made direct observations captured as field notes, and met afterwards to discuss strengths and weaknesses of the process and to organize their feedback.

The Evaluation Team also designed two participant online questionnaires (mid and post conference tools). They worked closely with the Leadership Team to refine and finalize these instruments to ensure they addressed the evaluation questions and could provide useful insight for the team. They administered these questionnaires using Qualtrics after the first full synchronous day at the conference (mid) and at the end of the last day of the conference (post).

Feedback and recommendations

The Evaluation Team provided their feedback to the Leadership Team at three stages in this project. First, throughout the conference development process, they offer their recommendations on the agenda, activities and selection criteria for attendees. Second, based on their meeting notes, observations and mid-conference questionnaire results, the Evaluation Team provided recommendations immediately after the first daylong meeting (April 15) so the team could make adjustments that were responsive to participant insights before the second daylong meeting (April 26). Finally, based on their direct observations, the Evaluation Team reported on their reflections associated with the two evaluation questions (provided below).

Findings

Evaluation Question 1: How successful was the team at diversifying participation and input from multiple perspectives? Did participants feel their voice and ideas were heard and valued throughout the process?

Based on questionnaire results and evaluator direct observations, the Evaluation Team found evidence that the Leadership Team was successful in diversifying conference inputs from multiple perspectives. Prior to and during the conference, the Leadership Team made significant efforts to ensure the event would contribute to diversity of participation and inclusion of thought. Evidence of these efforts include:

- The Leadership Team formed a culturally responsive and equitable evaluation and design (CREED) committee with three experienced practitioners (including Dr. Askew). Members of the CREED Committee were actively involved during Steering Committee meetings, and were regularly consulted outside of meetings to ensure the agenda-centered strategies advanced equitable and inclusive conference conversations.
- The Leadership Team collected and applied feedback from the Steering Committee in the recruitment and conference design. For example, Steering Committee members provided a recommended list of possible conference participants, and systematically reviewed each against diversity criteria. This recruitment process was documented to center goals of diversity, equity, and inclusion and to move beyond tokenism. It resulted in over a third of participants (11 of 26) of participants who represented or had deep experience in their work centering ethic/racial groups that have been historically marginalized in STEM. At the recommendation of Steering Committee members, the

Leadership Team also incorporated suggestions for working agreements that centered on relational values and use of multiple modes for participant engagement.

- The Leadership Team opened each synchronous session by explicitly stating their request for input that expanded the boundaries of the Framework. They also made time during whole group discussions for critical commentary around the relevance of the guiding constructs for diverse communities. Finally, during the second synchronous session, they describe modifications made thus far based on participants' insights and suggestions (e.g., grounding the Framework with case studies that drew from examples provided by conference participants).
- The Leadership Team allocated time to review and respond to feedback from both two committees (Steering and CREED) and the Evaluation Team's analysis of participant feedback after the first full day of the conference. They made extensive revisions in light of this guidance.
- Based on meeting notes and analysis of questionnaire responses, the Evaluation Team found evidence that the conference participants expressed their approval of the ways that the Leadership Team engaged and responded to diverse voices and perspectives during the conference. Productive dialogue transpired within and across the three types of attending professionals (researchers, evaluators and participants) as they shared their experiences, knowledge, and critical perspective on the Framework and its application. Second, participants appreciated the small group breakout sessions that were organized around their interests and supported by Steering Committee members serving as note-takers and online collaborative tools. Third, the participants praised the welcoming and inclusive tone set by the Leadership Team, highlighting the Team's emphasis on openness to constructive feedback, transparency, and reflection. Finally, during small and whole group discussions, participants demonstrated a strong willingness to revise the Framework, and, as noted, the Leadership Team reflected back what they heard from participants and how they planned to use this feedback. For example, participants problematized the use of the term "progressions" and observed that the ordering of the elements within a subject area could be interdependent and influenced by numerous contextual and cultural factors. Thereafter, the Leadership Team held this term lightly in conversation, and asked participants to explore new conceptualizations for each subject area.

While the Leadership Team made efforts to diversify the conference attendees, the Evaluation Team found evidence that diversifying the Leadership Team would increase the responsiveness of the Framework to address anti-racist and justice-centered ISE evaluation. For example, conversations frequently returned to whether the foundational underpinnings of the Framework would adequately disrupt harmful narratives and practices around STEM interest and STEM capital in ways that did not appear to be anticipated. Secondly, across more than one

small group breakout and during large group reflections, participants noted limitations of the current development process to advance cultural and contextual responsiveness. Participants made recommendations for additional input from evaluators (e.g., to help with its structure and presentation), as well other stakeholders include family and community members (e.g., to identify any missing aspects) to advance the Framework's ability to responsibly address issues of culture, equity, and inclusion.

Evaluation Question 2: How did the conference structure and process affect how participants felt about the Framework?

Based on meeting notes and questionnaire responses, the Evaluation Team found evidence that the conference process helped participants gain a better understanding of the Framework and its aim, and overall its potential to be relevant across diverse settings. Through their involvement with the Framework during the conference, conversation centered on important areas for change or reflection with regard to the Framework.

First, participants encouraged consideration of how outcome categories are measured, how they influenced each other and associated implications, how the outcomes are tied to practice, and how outcomes can be linked to case material. Second, while they acknowledged the complexity of the ISE field, participants recommended simplifying the Framework and providing explicit instructions (e.g., brief companion reference document), as well as concrete terms and definitions (e.g., categories, outcomes, impacts and audience). Third, participants proposed some shift in the perspective of the Framework--including thinking about it from an ecosystem perspective (e.g., connections to community health, well-being, formal education and other ecosystems components) and making equity more central to all of the outcome categories and/or incorporating diversity-equity-inclusion as its own outcome category. They noted it might help to prioritize these categories based on feedback from the field. Overall, the discussion reflected participants' recommendations to promote more flexibility in revisions of the Framework.

The participants also reflected on next steps for the Framework. They suggested applying for a NSF Research Coordination Network grant so researchers and practitioners could more thoroughly unpack the outcomes categories. They recommended documenting how the Framework influences an ISE program planning process (e.g., where does it pull users' attention, do the number of levels have an influence). They suggested consideration of how the Framework might influence the ISE field's definition of 'success' and how large organization goals could be focused around the outcome categories. Finally, they encouraged thinking about broadening applications of the Framework (e.g., learning goals of exhibits).

Overall, the Evaluation Team found evidence that, during the conference, there was not enough time to share complex insights or for small group discussions, and there could have been better transfer of conversations from small groups to the large group. Additionally, the

technology tools proved distracting at times (e.g., multiple people adding notes). Finally, as the conference development and implementation process demonstrated, attempts to establish a culturally affirming and inclusive environment must be done with intention and the support of a engaged CREED team; such efforts may be improved with support from a professional facilitator with experience in creating equitable processes and outcomes.

Final remarks

From participants' interactions and feedback during and after the conference, the Evaluation Team has identified several critical drivers, which can influence the continued development of the Framework and its potential to contribute to equitable impact across the field. First, for future work, the evidence supports the importance of recruiting and fully engaging a committed group of culturally diverse stakeholders, including researchers, evaluators, practitioners, and site-specific community members. This should include stakeholders and leadership team members who represented or had deep experience centering ethically, racially, and culturally diverse groups that have been historically marginalized in STEM. Together this group should enter discussions with a clear understanding of the Framework and the objective of the dialogue. Second, there should be clear communication throughout all team/stakeholder exchanges. This includes commitment to and communication of transparency of the Framework revision process, early articulation of roles and responsibilities, and facilitation of dialogues and associated work on the Framework that emphasizes inclusion (perhaps supported by a trained facilitator). Finally, the structure for these exchanges and the associated revision process should maintain a strong focus of equity, flexibility and adaptability. It should also make use of small group discussions and case material from the field. Additionally, there should be careful consideration of the complexity of possible supporting online tools for each reviewing task. And, of course, all of this work should consider the reality of the tumultuous world in which stakeholders and team live and work.

Appendix D. Proposed User Manual Outline with Notes

While it is beyond the scope of this white paper to create a user manual, the conference did produce rich, though partial, understanding of what the manual needs to cover. Some of those notes have evolved from the writing of this paper, others are direct quotes from participant discussions, comments, and feedback.

Overview of Purpose and Use.

What it is and what it isn't

Educational (learner) outcomes. Not visitor experience outcomes. (e.g., would you recommend to a friend, would you come back again).

"Some outcomes simply take longer, more sustained activities (like STEM capital, identity) … very hard to effect in a short session, so they might be an (overly) ambitious aim"

Individual participant educational/learning outcomes, not system-level outcomes (e.g., improve community recycling practices; create more employment opportunities)

The framework is meant to be used within, not between institutions.

Developed for use within the ISEE (Informal STEM Education Evaluation) System.

The framework has been developed to function within a larger system that involves a Voices and Choices planning process within which the framework nests. Within the system, the framework functions as one of four tools for unifying an understanding of an institution's full scope of educational planning. Following are a description of the ISEE system parts.

Voices and Choices Planning

The framework is designed to be used within a larger ISEE system, and specific to the effectiveness of choosing and using outcome categories and types within them, is that these choices are assumed to involve conscious decision making about the voices that have been involved in making them.

Within the ISEE system, the framework is intended to be used alongside of the "expectations for educational environment" unification tool. These expectations help to ensure conditions that lead to participant self-regulation and self-determination. Without that tool, it is important that along with selected outcomes, planning and feedback attend to participant sense of competence, relatedness, and autonomy (Deci et al., 1991; Wasserman, 2010) throughout their learning experience. Learning environments at each encounter with the learning experience -- welcome, orientation, exploration, navigation, support, and renewal -- need to meet learners' social, emotional, and physical needs (McLean & Pollock, 2011).

How the voices and choices planning tool is used at each of five stages of program planning through evaluation.

Manual. Voices & Choices. Planning (and implementation phases). Examples:

- Recognizing cultural factors that stimulate or prevent interest development and achievement in an area of interest such as conservation and the environment where there can be specific cultural orientations toward engaging with the environment.
- Educators make a difference. Their interest in the content affects their interest in supporting others (visitors/participants) to develop an interest in it.

- Consider how external factors such as lack of resources influence the development of interest or the depth of any other type of learning.
- Considering that a direct way to structure learning experiences in a way participants make a meaningful connection to the content is to include elements that are self-related. Representative voices in the planning can help make these connections more accessible. Voices during the implementation means designing activities allow participants to be "involved in, for example, identifying what they already know, how information about x is like something in their lives, etc. If there is some way to sustain the questioning, or activity, so that they are supported to explain how they know and why they thin what they do this could be useful. Another way to provide self-related content is to anchor the content to be learned using culturally-relevant, or personalized, contexts- that would allow the visitor/ participant to make links between their own lives and that of the content of the exhibit/program." (Ann Renninger, interest discussion board.)

Framework use

As a component of the ISEE system;

As a guide for both creation and co-creation of program outcomes and activities that support them;

As a guide for program evaluation

Evaluation capacity building

Not from institution to institution but within institutions.

Funder awareness tool

To encourage definition of terms at institution and program level

How the Framework can support ISI commitment to Addressing Social Justice Issues

Consider contribution analysis. Support for the framework needs to include how to construct learning outcomes that lead to system impact. Individual learning contributions can involve skills, behavior, attitudes, etc.

Use of the framework outside of the ISEE system.

Caveat.

Without the system, use of the framework needs to be accompanied by (1) a way to include voices of key stakeholders and (2) inclusion of self-regulation/self-determination factors (which the system does via the educational environment expectations tool.)

The Framework Description

Category overview,

Outcome Types within Categories-

The Outcome Categories

The outcome categories (in the framework rows) derive from two sources. In 2008,participants of a National Science Foundation STEM Education Evaluation Activities project, workshop participants

identified six "impact categories" as part of framework for Evaluating impacts of Informal Science Education Projects (Allen et al., 2008). In 2009 the Committee on Learning Science in Informal Environments offered "six interweaving strands that describe goals and practices of science learning" noting that "while these strands reflect conceptualizations developed in research, as a set they have not been systematically applied and analyzed."(Bell, Lewenstein, Shouse, & Feder, 2009, p.43). As shown in Table 5, these strands can be mapped to the Allen et al. (2008) categories.

Table 5. Two learning outcome frameworks compared.

Impact Categories (Allen et al, 2008)	Informal Science Learning Strands (Bell et al, 2000)
Engagement or interest Engagement/interest in a particular scientific topic, concept, phenomena, theory, or careers	Strand1: Experience excitement, interest, and motivation to learn about phenomena in the natural and physical world.
Awareness, knowledge or understanding	Strand2: Come to generate, understand, remember, and use concepts, explanations, arguments, models, and facts related to science.
Skills—procedural aspects of knowing	Strand3: Manipulate, test, explore, predict, question, observe, and make sense of the natural and physical world.
Attitude- relatively stable, more intractable constructs (e.g. empathy for animals; appreciation of scientists in society; or attitude toward stem-cell research	Strand4: Reflect on science as a way of knowing; on processes, concepts, and institutions of science; and on their own process of learning about phenomena.
Behavior. Action is the desired outcome.	Strand 5: Participate in scientific activities and learning practices with others, using scientific language and tools.
Other. Project specific.	Strand6: Think about themselves as science learners and develop an identity as someone who knows about, uses, and sometimes contributes to science.

This the strategic outcome framework categories have expanded on these sources in two ways. First we have extended the definitions beyond "Science" to include all of STEM: Technology, Engineering, and Math. Second, as pilot projects began working with the framework we recognized the need to expand the strands to include STEM Capital; Career Activation & Development; and 21st Century Skills (as separate from STEM skills). The definitions provided in each of the descriptive sections below emerge largely from these two sources.

Category	Definition
Interest	The state of wanting to be with, know or learn about something or someone.
Attitude	Personal reflections on concepts, processes, institutions, phenomena,
Knowledge	Memory or understanding of facts, concepts, explanations, arguments, models
STEM Skill	Procedural aspects of knowing. Includes Manipulate, test, explore, predict, question, observe.

21st century skill and social-emotional development/learning	Self-concept and emotion knowledge; social awareness; emotion regulation; goal management; school work; relationship; decision making; critical thinking; creative thinking; et al.
Behavior	To be defined
STEM Capital	What science you know, who you know, how you think, what you do; what may help transform positive attitudes and interest in STEM into actions that make future science participation more likely
Career Path	Used as a measure of long-term impact of STEM education and engagement and as a measure for scientific identity development

- Differentiate interest from attitude. Select interest when the ultimate aim could be a long-term relationship to a particular topic. Select attitude when the program is specifically aimed toward, or the primary outcome specifically requires, a qualitative affective relationship and commitment to a topic

Interest.

Definition

- Needs to include examples of how interest is defined, operationalized, as well as types of programming that triggers it.
- "If a person has interest, they are motivated and engaged (and note, it doesn't work the other way- you can be motivated, but not interested, and you can be mindlessly engaged, and not interested). One of the things that is complicated about out-of-school settings is that they their facilitators/and even their visitors expect/assume that they promote interest, but this assessment isn't always conducted, or it isn't conducted in ways that end up being useful to the design team.
- Include explanation of curiosity as it fits in the definition of interest. such as these belong in these sections of the user manual.

Planning strategies

Ways to structure activities/programs so that situational interest leads to stable interest & goal setting. How to get regular feedback about situational interest and self-regulation (i.e., choice making that is free from tension, pressure, or ambiguity). This step would involve voices and choices planning process at planning and implementation phases.

Evaluation

- Data Collection Methodology
- Outcome Indicator statements and instruments

Interdependence

"Interest and its development is important to developing (or promoting others to develop) an identity as a person who thinks that they can and continues to pursue STEM, for example. Interest does not exist in a vacuum. It has a relationship with other motivational (as well as instructional) variables.

Include examples of how interest supports other outcomes categories, e.g., behavior change.

Knowledge

Specifying outcomes

Include guide for specifying outcomes within outcome types..

Interdependence with other categories

Knowledge can support attitude and vice -versa

STEM Skill/Practice

Interdependence with other categories

Manual. Categories. Skill/Practice. Knots. Using Practice as a building attitude toward learning science.

Strategies

Consider practice as a strategy for achieving other outcomes.
How do you achieve “practice” virtually?

Examples

Hybrid practices: how youth "merge" their personal and community knowledge with STEM knowledge to build new and deeper understandings of something. [DW here, what is the outcome? Building new and deeper understanding of something? (Knowledge level 3,4, or 5)] Consider, for example, how a group of 6th grade class was studying statistics and the teacher purposefully used the context of evaluating playground safety to teach about environmental statistics. He hoped that this would "create spaces" for students to leverage on their knowledge of how playgrounds work, and the issues that matter in playgrounds.

Students [who] participated used their knowledge to not only deepen their understanding of statistics [DW: outcome = knowledge/understanding of statistics—did the knowledge/understanding involve developing a skill?] but also to shift the collective outcomes for the whole class, including, arguing for a “real” evaluation that makes a difference, changing evaluation criteria to match more realistic concerns, not participating because playgrounds are for little kids and so on [DW: this outcome is trickier. This may be an applied use of STEM practice that is indeed different than a skill trajectory. Maybe we need both. I’m also interested your point about “changing evaluation criteria to match more realistic concerns.” So maybe practice outcomes have to do with “relevance” and “value” as with attitude. [Angie from STEM Skills discussion]

Selecting Categories, Identifying Outcome Types, and Specifying Outcomes.

- Support for the framework needs to include how to construct learning outcomes that lead to system impact. Guides for framework use or future use of the framework might consist of examples of which kinds of learning outcomes can weave together to contribute best to longer term and larger system impact. Similarly, future framework use might function as ways for educators to structure consideration of how to best program to achieve these impacts.

Understanding outcomes as outcome types. Within which outcomes are specified

Knots: Primary, supporting, and “reach” outcomes.

Outcome selection. Note that within categories outcome types can be used as a “flat” menu or they can be sequenced in a way that a learner can locate themselves along a continuum.

Manual. Selecting Outcomes. Primary outcome: Ask first, what is the primary intention of this program. Then ask, what outcomes will support arriving at that primary outcome

Manual. Category. Selection. Example of choosing between two categories. STEM "practice" vs. behavior. The difference would be specifically related to the program and how it identifies itself. Is the program about conservation behaviors as STEM practice, i.e., awareness of STEM as a pathway to knowledge? Or as a behavior that addresses a social/ecological/environmental issue. Or both?

Examples.

Knowledge supports skills and skills support knowledge.
Reflection. Categories. Relationships. There may be a supporting relationship from situational interest (the way a person participates) to deeper interest (motivation to re-engage) to "identity"

Where does this outcome intention belong?: Case examples of questions to ask in order to select categories, arrange outcome type, and specify outcomes.

- Social justice issues and system change
- Social-emotional development (can it be part of 21st century skills or does it need its own category?)
- Self-efficacy, self-confidence, and sense of competence
- Self-confidence Depending on a program's focal point for self-confidence, it could involve attitude. For example self-confidence in ability to learn about science in school follows the attitude depth levels. On the other hand self-confidence alone doesn't (i.e., most people already have a positive association to self-confidence and it has personal relevance and pd value) etc.

STEM Identity

Reflection. STEM identity. The framework recognizes the complexity of this term. STEM identity could be any combination of Attitude 4,5,6 (to do: define STEM identity in terms of potential framework learning outcomes

- Building STEM Community
- STEM literacy
- Beliefs and Values
- Empathy
- Constructivist, serendipitous learning experiences.
- Applying STEM to address social and other non-STEM issues.
- Maybe there is also a 21st century skill of "recognizing and communicating what I already know.)

Glossary

Include:

Category
Eight category types
Category scope
Outcome type
Outcome specification

Indicator statement
Primary outcome type
Supporting outcome type