

# FRAMEWORK FOR SUMMATIVE EVALUATION: EVIDENCE FOR ITS VIABILITY

Alice C. Fu, Lisa Peterson, Archana Kannan, Richard J. Shavelson, & Amy Kurpius  
SK Partners, LLC

## Project Overview

The purpose of the work presented, in broad terms, is to contribute to strengthening evaluations in informal science education (ISE) and building capacity for doing so. We focus here on summative evaluations, which have been called upon to inform decision-making and practice, contribute knowledge to the field, and help make the case for the value of informal experiences. To serve these functions, summative evaluation must be high quality. Our framework synthesizes key elements of high-quality summative evaluation—a guiding lens for planning or conducting current or future summative evaluations and also in reviewing or using completed evaluations. We highlight findings from research conducted on the viability of the framework, i.e., whether it adequately defines criteria for quality and how these criteria might be used in practical ways.

### Evaluation Uses

**Uses:** How are findings and recommendations framed to support stakeholders and the field in using the evaluation? What is the evaluation's impact?

**Communication:** Are communications appropriate for the audience(s), clear, digestible, and timely?

**Stakeholders:** Who are they? What are their needs?

*In our case study of a summative evaluation at a children's museum, stakeholders used the evaluation to inform remediation, support the future use of multi-lingual exhibit labels at their museum and elsewhere, and influence institutional initiatives. "They [at the museum] need and expect and want their evaluation reports to inform future work rather than just tell them how they did" (Interview, 7/23/13).*

*Our review of reports found that actual uses of an evaluation are typically undocumented, as these uses usually occur after report writing and dissemination.*

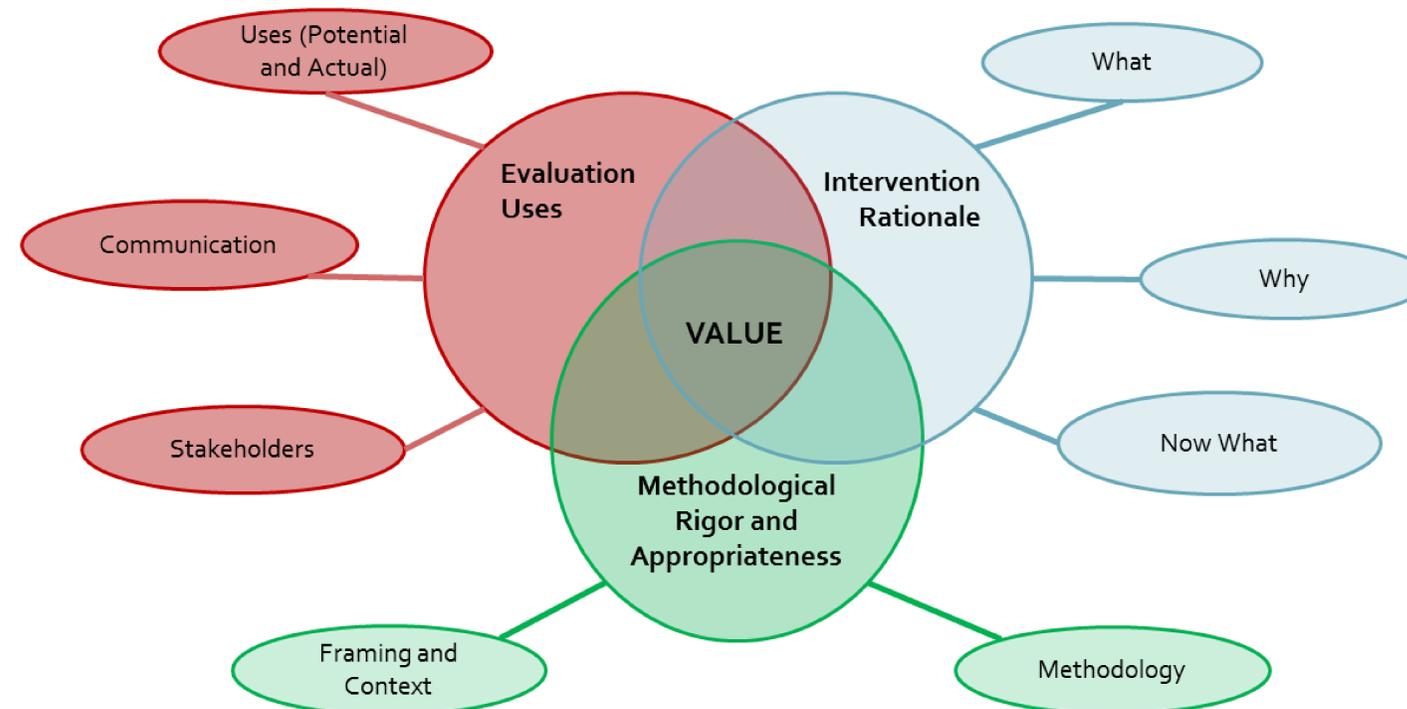
### Methodological Rigor and Appropriateness

**Framing and Context:** Are scope and purpose appropriate for the resources available and the intervention's maturity and stability? Is the evaluation sensitive and authentic to the context?

**Methodology:** To what extent are questions, design, methods, interpretations, and conclusions tightly and explicitly linked? Is fidelity of implementation examined? Are generalizations appropriate? How are unexpected findings taken into account?

*Rigor hinges on suitability of the study for the evaluation purpose and questions, not on particular designs or methods. Yet, our review of summative reports found heavy reliance on a few designs and methods: 83% (N=30) of 36 evaluations used non-experimental designs, 50% (N=18) used quasi-experiments, and 3% (N=1) used a randomized experiment; 75% (N=27) of reports used interviews, and 67% (N=24) used surveys. (Percentages do not total 100, as multiple designs and methods may be used per report.)*

*In interviews, we heard about the challenge of balancing methodological rigor and contextual appropriateness. "...you have to make some compromises...remaining as true as possible to the free choice nature, and if you are distorting it too much, your results are just not going to be worth as much...that is something that we [as evaluators] have to think about all the time" (Interview, 8/26/13).*



### Intervention Rationale

**What:** What is the substance of the intervention (program, exhibition, etc.)? What do participants experience?

**Why:** What is the underlying rationale and justification for development? What are reasonable alternative interventions?

**Now What:** Is a summative evaluation warranted at this time? What bears closer examination?

*Our review of reports found that, although evaluators may develop deep understanding of the intervention and its rationale, little of this is typically reported.*

*Kearny's (2009) report of a summative evaluation at Islandwood (Bainbridge Island, WA) is a rare example. Using a logic model, extensive references, and data on stakeholders' priorities, the program rationale and basis for evaluation questions are clearly and explicitly presented.*

### Value

Summary judgments or reflections on the intervention's value should be reasonable and justified, drawing upon all framework dimensions—information about the intervention and its rationale, empirical evidence from the evaluation, and an understanding of what is useful for stakeholders.

*"...I have said [to evaluators], 'Where is your sense of, how well did we do?' They would say 'That's not my job.' I really don't agree with that ... We need to be saying, 'Here are the strengths, here are the failures, and here are the interesting issues'" (Interview, July 10, 2013).*

## Project Methods

We developed an initial framework based on a literature review and then refined it iteratively over the course of our research. Activities included: (1) **Case study:** We selected a summative evaluation report that scored highly on criteria from our framework. We conducted a case study of this evaluation of an exhibition at a children's museum, interviewing 7 key players to understand how it was designed, conducted, and used. (2) **Interviews:** We conducted interviews with 8 ISE leaders and asked for their perspectives on pressing evaluation issues and the future of the field. We also interviewed individuals from 3 research/evaluation studies that employed experimental designs in informal settings. (3) **Review of reports:** We reviewed all 2012 summative evaluation reports posted on informalscience.org (N=36) to gauge commonly-used study designs and methods. We also used criteria from our framework to review other reports gathered from purposive snowball sampling, focusing on those recommended as exemplary (by colleagues, team members, or citations in other works). **Qualitative data** (i.e., interviews, reviews of reports) were analyzed according to the main dimensions of our framework, as well as for other, emergent themes. **Quantitative data** (i.e., results from coding the 2012 informalscience.org summative evaluation reports) were analyzed for frequencies of reports in the sample with various characteristics such as type of study design and methods/instruments used.