Enhancing Evaluation in Informal Science Education: Review of Summative Evaluation Reports

Coding Sheet and Framework Worksheet

Contents

Introduction

For our review of summative evaluation reports in informal science education (ISE), we took a two-pronged approach.

First, we examined methodological characteristics of summative evaluations in informal science education (ISE), asking: What are the major types of designs used in summative evaluations, and what kinds of questions can they answer? What are the types of data collection methods and measures used, and how many are self-reports or direct measures? We analyzed all of the summative evaluation reports published in 2012 on the Center for Advancement of Informal Science Education's (CAISE) website—http://www.informalscience.org, a central online resource of the ISE community. We searched the informalscience.org database using two parameters: (a) Type of resource: summative evaluation reports, and (b) Year: 2012. Our search yielded 36 results, and we reviewed all 36 summative reports. The team developed a **coding sheet** to systematize their efforts. Five reports were independently reviewed by two researchers to establish interrater agreement (97% interrater agreement).

Second, we examined more closely a separate, smaller sample of summative evaluation reports. Our Framework for Summative Evaluation suggests some essential features of high-quality evaluations, and we developed a **framework worksheet** for systematically considering these features in evaluation reports. We propose that this application of the quality framework may be of interest to those who must review a variety of reports and/or outline in the first place what should be included in them. Additional reliability and validity evidence is needed to determine whether this approach can be used feasibly and broadly.

This document includes:

- A coding sheet for "quick" coding of basic descriptive and methodological features of summative evaluation reports;
- A codebook that defines terms used in the coding sheet; and
- A framework worksheet for in-depth analysis of an evaluation report according to the dimensions of the Framework for Summative Evaluation.

The coding sheet and framework worksheet can be used separately or together depending on purpose and resources. The former facilitates a faster review of reports and is focused on evaluation methods, while the latter guides a more comprehensive analysis of reports that includes questions about the rationale behind the intervention being evaluated, the rigor and contextual appropriateness of the evaluation design and methods, and uses of the evaluation.

Coding Sheet for ISE Summative Evaluation Reports

Title of Report:
Date:
Authors:
Name of Intervention:
	Name of Institution:
	   Evaluated at:
	   Developed by:
	Name of Evaluator:

Reviewer Initials:	

Pages	Type of Intervention	Institution (Evaluated at)	Type of Evaluator

___(exec. summary)
___(body)
___(appendices)
___(other)
___(total)
 	
References
 	
____ (number)
	□ Designed Settings (exhibits)
	□ Programs/Events/Activities
	  □ schools (e.g., field trips)
	  □ children and teens (non-school)
	  □ adults/general community
	  □ professionals
	□ Cross-Cutting
	  □ Mass media
	  □ Games & Learning technologies
	□ Other _____
		□ ISI
		  □ aquariums and zoos
		  □ arboretums/botanical gardens
		  □ museums and science centers (science-rich)
		  □ children's museums
		  □ nature centers (incl. interpretive centers of parks, preserves, etc.)
		□ General (non-science-rich) museums
		□ Community/Youth Organizations
		  □ science-rich
		  □ non-science-rich
		□ Other _____
			□ Internal
			□ External
			□ Unsure
			
			Type of Methods
			
			□ Qualitative
			□ Quantitative

Method Code	Method (see key below)	Timing					Sample Size	Comparison or Control Group	Random assignment (if control group)	Instruments included in report?	Notes
		Pre	Retro Pre	During	Post	Delayed Post					

Methods Key: 1-Survey (self), 2-Survey (other), 3-Interview, 4-Focus Group, 5-Observation (tracking & timing), 6-Observation (other), 7-Video and Audio, 8-Test, 9-Mapping Exercise, 10-Performance Assessment, 11-Other. [See Codebook for more information.]

Codebook for Coding ISE Summative Evaluation Reports

Bibliographic and Descriptive Information:

- **Title of Report:** Title of the report as in the title page
- **Date:** Date of report – year of publication; month and date if available
- **Authors:** Names of the people who wrote the report. Usually found in title page. If author not mentioned, fill in "not available."
- **Name of Intervention:** Name of the program, exhibit, etc. being evaluated in *this* summative evaluation. Usually found in introduction or executive summary sections.
- **Name of Institution**
  - **Developed by:** Name(s) of all collaborating institutions that developed this intervention
  - **Evaluated at:** Name(s) of all institutions where the intervention was evaluated in *this* particular evaluation
- **Name of Evaluator:** Name of the individual, organization, or company who did the evaluation. Usually found on the title page(s) and/or introduction section.

Pages:

- (exec. summary): Total number of pages in executive summary (plus abstract, if available)
- (body): Number of pages in body of the report (include title page of body and references)
- (appendices): Number of pages of appendices (include title page for appendices)
- (other): Number of pages excluding summary, body, and appendices
- (total): Total number of pages in the report. The numbers above should add up to this total.

References (number): Total number of citations to other literature in the report. This includes references to all research articles, reports, websites, and books. If there is a reference section, count the number of citations in it. Also check for other references and citations within the body of the report and in footnotes. Count both and record the total number here (ensure that there are no duplicates).

Type of Intervention [Select ALL that apply]:

- **Designed Setting:** Primarily exhibits and exhibitions in museums, science centers, nature centers, botanical gardens, zoos, aquariums, libraries, etc. "Artifacts, media, and signage are primarily used to guide the learner's experience. While these environments are structured by institutions, the nature of the learner's interaction with the environment is often determined

by the individual. Learners also have significant choice in setting their own learning agenda by choosing to attend to only exhibits or aspects of exhibits that align with their interests. Typically, learners' engagement is short-term and sporadic in the setting" (NRC, 2009, p. 48).[1] [NOTE: Do not check "Designed Setting" if the intervention is a structured program that also involves learners visiting exhibits in a museum (such as a school program, after-school program, a family/adult program). See Programs/Events/Activities for coding various types of structured and pre-planned programs and learning experiences.]

- **Programs/Events/Activities:** Includes all structured and pre-planned learning experiences. "Often program content includes a formal curriculum that is organized and designed to address the concerns of the sponsoring institutions. Experiences in these environments are typically guided and monitored by a trained facilitator and include opportunities for hands-on, collaborative experiences. The time scale of these learning experiences ranges from being sustained, long-term programs with in-depth engagement to brief, targeted, short-term programs" (NRC 2009, p. 48). We have categorized programs into four subtypes: programs for schools, children and teens, adults/community, and professional. If the specific subtype is unknown, just select programs.
    - **Schools:** Programs for elementary, middle and high school students *during school hours.* These are typically classroom field trips to museums, science centers, aquariums, zoos, etc. School field trips exist on a continuum - they can be highly structured experiences that include a formal curriculum and pre-planned activities, or they may involve kids just wandering around the museum exploring on their own. ALL types of school field trips are included here, even if they involve kids just wandering around. (The reason is that a field trip is still a structured activity: kids are coming to the museum because it was scheduled by a teacher; they arrive at a designated time, have a certain amount of time to explore, and leave at a designated time. This is different from a non-school trip visit where those things would be more free-choice.)
    - **Children and teens:** These include programs for school kids typically *outside of formal school time*. Examples are after-school programs, out-of-school time programs, weekend programs, summer programs, residential programs, out-of-school time high-school programs, high school internships in an informal learning environment. "Experiences in these environments are typically guided and monitored by a trained facilitator and include opportunities for hands-on, collaborative experiences. The time scale of these learning experiences ranges from sustained, long-term programs with in-depth engagement to brief, targeted, short-term programs" (NRC, 2009, p. 48).

---

[1] National Research Council [NRC]. (2009). *Learning science in informal environments: People, places, and pursuits.* Committee on Learning Science in Informal Environments. P. Bell, B. Lewenstein, A. W. Shouse, & M. A. Feder (Eds.). Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=12190

- **Adults/general community:** Programs targeted at general public - typically adults and families - fall into this category. These include family programs, community programs, hobby clubs, science cafes, volunteer programs for adults, lecture series, and so forth.
  - **Professional:** Professional development programs for teachers, formal and informal educators, graduate students, and scientists.
- **Cross-Cutting:** Check this only if the focus of the evaluation includes cross-cutting features that can be applied in either designed settings or programs; these cross-cutting features include educational broadcast media, games, and learning technologies. Do not check this if media or technology is just a part of an exhibit or program.
  - **Mass media:** Television documentaries, nature and science films, printed news stories in newspapers/ magazines, radio programs, national broadcasts, internet media (email, websites, blogs, internet-based radio and videos).
  - **Games and learning technologies:** Online games, video games, in-museum games, and technology-based interventions.
- **Other:** For any intervention that does not fit into the above categories, select "Other" and describe it.

Type of Evaluator [Select ALL that apply]:

- **External:** Evaluator who is hired or contracted from outside the institution where the intervention was developed. The evaluator is generally someone who was (or is) not directly involved in the development or operation of the intervention. If the evaluator's institution is different from the institution that developed the intervention, then it usually means external.
- **Internal:** Evaluator is from within the institution where the intervention was developed. An internal evaluator may be someone from the intervention development team or may be part of a research and evaluation department (or similar) within the same institution.
- **Unsure:** If it is not evident from the report whether it is an internal or external evaluator, then check unsure.

Type of evaluator may or may not be explicitly stated in the report. When not stated, it can be inferred from the information reported about authors, name of evaluator's organization, name of the (internal) department. Information about type of evaluator is typically found in the title page(s), suggested citation, executive summary, or introduction section.

Type of Methods [Select ALL that apply]:

- **Quantitative:** Any method that makes use of numbers - statistics, frequencies, percentages - to describe data.

- **Qualitative:** Any method that uses verbal descriptions - quotes, words, narratives - to describe data.

The actual terminology "qualitative" or "quantitative" may or may not be used in the report to describe the methods; it can be inferred from the methods and/or the results sections of the report.

Type of Institution (evaluated at) [Select ALL that apply]:

Select the type(s) of institution(s) *where the intervention is evaluated* in this evaluation. For some interventions (such as traveling exhibits, after-school programs, etc.), the report may refer to multiple institutions where the intervention was evaluated; in such cases, select all that apply.

- **ISI:** Informal Science Institution. This category includes all institutions with a focus on STEM learning. Pick appropriate subtype from the list:
    - **Aquariums and zoos**, e.g., Monterey Bay Aquarium, Oakland Zoo.
    - **Arboretums and botanical gardens**, e.g., San Francisco Botanical Gardens.
    - **Museums and science centers (science-rich)** (including natural history museums), e.g., Exploratorium, Oregon Museum of Science and Industry, California Science Center, Smithsonian Institution National Museum of Natural History.
    - **Children's museums** e.g., Children's Discovery Museum, Chicago Children's Museum.
    - **Nature centers** (including interpretive centers of national, state, and local parks; nature preserves, etc.), e.g., Grand Canyon National Park.
- **General Museums**: Museums not focused on STEM; includes art museums, history museums and others, e.g., Crystal Bridges Museum of American Art.
- **Youth/Community Organizations**: Includes organizations that run after-school programs or educational programs and events for the public, e.g., YMCA.
    - If the organization is focused on science learning, then check science-rich.
    - If not, check non-science-rich.
- **Other**: Any institution that does not fit into the above categories. Use keywords to describe the institution type (such as media production companies, universities, libraries), e.g., BBC, NOVA.

Methods Table:

Every method in the report should have its own row in the table. For each method, indicate its numeric code (1-11) in the first column, and write a brief description of it in the second column. If the same type of method is used multiple times but with different samples or in different sub-studies, then create a new row for each method. For example, a single evaluation might include

interviews with visitors and interviews with staff; each should occupy a different row in the methods table.

- **1-Survey (self):** Self-administered survey. Participants fill out the survey on their own. Can be online, computer, or paper-based. Surveys typically have a set of structured questions or items and can be open-ended, closed-ended, or both.
- **2-Survey (other):** Survey administered by the researcher, evaluator, or other project staff. The evaluator follows a survey protocol, asks questions to the participant and fills out their responses. This type of survey can be administered in-person or over the phone. Survey questions can be open- and/or closed-ended.
- **3-Interview:** Participants are interviewed by the researcher, evaluator, or other project staff. Interviews may be semi-structured or unstructured, and the interviewer may or may not follow a protocol. [NOTE**:** The line between an *evaluator-administered survey* and *interview* can be fuzzy and we try to more clearly define this "boundary" for our coding purposes. A highly-structured interview where the evaluator deviates little from the set protocol is essentially an evaluator-administered survey and is categorized as Survey (other). Semi-structured and unstructured interviews that allow room for in-depth and unpredicted exploration of topics are categorized as Interviews. If the interview or survey instrument is not available, then we code it according to whatever is said in the report, i.e., if the report calls it an interview, then we code it as an interview; if the report says it's a survey, then we code it as a survey.]
- **4-Focus Group:** A research method in which a group of people participate in a guided discussion about an intervention (exhibit or a program). Questions are asked in an interactive group setting where participants are free to talk with other group members.
- **5-Observation (tracking & timing):** Observation where visitors are followed/tracked unobtrusively through the museum or an exhibition, and their behaviors are noted (often using a floor plan or an app). The data collector tracks the visitors and times them for how long they spend at various components and areas of an exhibition. This method can determine how much time the average visitor spends in an exhibition. "Hot spots" and "cold spots" can be identified, showing where visitors spend the most or least amount of time as they move through the exhibition. Usually used in museums, zoos, and aquariums.
- **6-Observation (other):** Observation of visitor or participant behavior that is not timing and tracking. Such observations typically have an overarching question about some particular visitor experience or behavior. They do not systematically track visitors step-by-step.
- **7-Video and Audio:** Video or audio recording of visitor or participant behaviors and conversations in an exhibition or program.
- **8-Test:** Use of multiple-choice items, open-ended questions, typically to measure visitor learning, knowledge, understanding, or awareness.
- **9-Mapping Exercise:** Visitors or participants are asked to construct a map or diagram, e.g., concept maps, personal meaning mapping.

- **10-Performance Assessment:** "A test in which the test taker actually demonstrates the skills the test is intended to measure by doing real-world tasks that require those skills, rather than by answering questions asking how to do them" (Retrieved from http://www.ets.org/understanding_testing/glossary/, October 7, 2013).
- **11-Other:** If the method used does not fit into any of the above categories, pick "Other" and describe the method.

For each method/row, fill in the following details, as appropriate: timing, sample size, use of comparison/control groups, random assignment, and whether or not instruments are shared in the report.

Timing:

- **Pre:** Measure administered to participants before they experience the intervention
- **Retro-Pre:** Measure administered to participants *after* they experience the intervention but asking them to reflect on what their perceptions, beliefs, knowledge, behaviors, attitudes, etc. were *before* the intervention.
- **During:** Measure administered during the intervention.
- **Post:** Measure administered to participants after they experience the intervention. This can be immediately after the intervention (e.g., an exit interview that is done as participants are walking out of an exhibition), or it may be done a few days or weeks after the intervention experience. The intention is to get a measure of participant outcomes post experience.
- **Delayed Post:** Measure administered to participants, typically weeks or months after they experience the intervention. It is usually done with the intention of either measuring outcomes after a specific time period and/or comparing pre or post-test outcomes with delayed-post outcomes to see if there are any long-term changes.

Sample Size: Indicate the number of participants in the sample, if this information is available. If information is not available, write N/A.

Comparison/Control Groups and Random Assignment:

- **Comparison/Control Group:** Yes or No. For a single method, if the treatment group is compared to a control or comparison group, write "yes" in the comparison/control group column. If there is no comparison group or control group, then write "no." Do not include comparisons within the treatment group, such as gender, age, ethnicity etc. If there is no comparison group within this study, but comparisons are made with standards in the field or other studies, write "No" and include notes explaining the type of comparison made.

- **Random Assignment:** Yes or No. If there is a control group, and participants are randomly assigned to treatment and control groups, then code as "Yes" in the random assignment column. If not, code as "no."

Instruments Included: Yes or No. Check whether the full instrument is included in the report. Instruments are usually found in the appendices, but sometimes they may be in the body of the report.

Framework Worksheet for ISE Summative Evaluation Reports

This worksheet has two parts: (a) a Table of Guiding Questions, and (b) a Summary of Review. For in-depth review of reports, we used the following general process:

1. Complete the "quick" Coding Sheet (see earlier in this document).
2. Complete the Table of Guiding Questions. This table walks the reviewer through a detailed set of questions, drawn from the main dimensions of the Framework for Summative Evaluation. The reviewer records what worked well and what could have been done differently in the evaluation report.
3. Prepare a Summary of Review using the completed Coding Sheet and Table of Guiding Questions.

Framework Worksheet: Table of Guiding Questions

Framework Dimension	What Worked Well	What Could Have Been Done Differently
Intervention Rationale: What What is the substance of the intervention? Who are the participants and what are they doing? What does the intervention look like when implemented?		
Intervention Rationale: Why What is the underlying rationale and justification for development? Why was the intervention done in the first place? Why was the intervention designed and developed in this way? Is there theoretical, empirical, and/or practical grounding or evidence to support its design and development? Have moral (i.e., "do no harm") considerations been taken into account?		
Intervention Rationale: Now What Is a summative evaluation warranted at this time? What questions, program elements, links in logic, or assumptions bear closer examination?		

Framework Dimension	What Worked Well	What Could Have Been Done Differently
Methodological Rigor and Appropriateness: Methodology Is there methodological rigor? Several elements should be explicitly and tightly-linked in the evaluation's "chain of reasoning": questions, study design, data collection instruments and methods, analysis, results, interpretations, and conclusions. Are measures and methods fully accessible for others to examine, replicate, adapt, or otherwise learn from? Is fidelity of implementation examined? Are generalizations to other people or contexts appropriate, with limitations spelled out? Are unexpected findings taken into account?		
Methodological Rigor and Appropriateness: Framing & Context Is the scope and purpose of the evaluation appropriate for the resources available and the "life stage" (maturity and stability) of the intervention? Is the evaluation study contextually appropriate? Do the design and methods take into account factors such as participants' expectations for an informal experience, degree of choice, engagement, and maintenance of a positive experience?		

Framework Dimension	What Worked Well	What Could Have Been Done Differently
Evaluation Uses: Stakeholders Does the evaluation project specify its intended stakeholders and their needs? What are the intended uses of the study?		
Evaluation Uses: Communication Is the mode of communication appropriate for the audience(s)? Is it transparent, digestible, and timely? Is it sufficiently detailed without being unreasonably long? Does it include conventions such as an executive summary, clear headings, appropriate visualizations and a logical flow if ideas? Does the communication / dissemination strategy seem to fit stakeholder needs?		
Evaluation Uses: Uses Does the evaluation usefully frame the findings for the intended stakeholders and purpose? Are recommendations appropriately grounded in evidence, tied back to the study questions, and actionable? Once the evaluation study is completed and disseminated, what is its impact? To which stakeholders is it ultimately useful and in what ways?		

Framework Worksheet: Summary of Review

Report Citation:	
Total # of Pages	[Total number of all pages in the report]
Name of Evaluator	[Name of the individual, organization, or company who did the evaluation.]
Type of Evaluator	[Was the evaluator internal or external? See Coding Sheet.]
Name of Institution	[Where was the evaluation conducted?]
Type of Institution	[What type of institution was the evaluation conducted at? See Coding Sheet.]
Reviewer Initials	[Initials of the person who reviewed the report and prepared this summary]

The Intervention	
Name of Intervention	[Name of the program, exhibition, etc. being evaluated]
Type of Intervention	[Type of intervention being evaluated. See Coding Sheet.]
Description of the Intervention	[Briefly describe the intervention being evaluated.]

The Evaluation	
Overview	[Briefly summarize or copy/paste an overview of the evaluation purpose or goals.]
Outcomes	[List the types of outcomes that were evaluated. If at the level of individual participants, types of outcomes might include: engagement, cognitive, intra-personal, inter-personal/social, behaviors, and so on.]
Methods	[Qualitative and/or quantitative? See Coding Sheet.]
Design	[Briefly describe the study design, i.e., experimental, quasi-experimental, pre-experimental, case study, ethnography, other.]
Comparisons	[List any types of comparisons found in the report, i.e., traditional control group, matched control groups, non-equivalent comparison group, comparisons over time, compared to other projects/studies, compared to standards/benchmarks, compared to goals.]
Findings/Claims	[Briefly summarize or copy/paste the evaluation's main findings and claims.]

Reviewing the Report
[Complete the following based on key findings from the completed Table of Guiding Questions.]

Intervention Rationale	
What Worked Well	**What Could Have Been Done Differently**

Methodological Rigor and Appropriateness	
What Worked Well	**What Could Have Been Done Differently**

Evaluation Uses	
What Worked Well	**What Could Have Been Done Differently**

Example of Coding Sheet and Framework Worksheet: Review of Secrets of Circles Summative Evaluation Report

In the following pages, we provide an example of a completed "Coding Sheet for ISE Summative Evaluation Reports" and a completed "Framework Worksheet: Summary of Review."

We reviewed Allen's (2007) report on her summative evaluation of the *Secrets of Circles* exhibition at the Children's Discovery Museum of San Jose. Our review identified several exemplary features of the evaluation report. In search of additional information about and perspectives on the *Secrets of Circles* summative evaluation, we conducted a case study of the project. More information can be found at: http://www.informalscienceevaluation.org/case-studies-of-exemplary-evaluations.html.

Example of Coding Sheet for ISE Summative Evaluation Reports

Title of Report: *Secrets of Circles* Summative Evaluation Report **Date:** October, 2007 **Authors**: Sue Allen **Name of Intervention:** Secrets of Circles	**Name of Institution:** **Evaluated at:** Children's Discovery Museum of San Jose (CDM), Strong National Museum of Play, Ecotarium **Developed by:** CDM **Name of Evaluator:** Sue Allen, Allen & Associates
Reviewer Initials: [SK Partners]	

Pages	**Type of Intervention**	**Institution (Evaluated at)**	**Type of Evaluator**

6   (exec. summary)
 124 (body)
 39 (appendices)
 4   (other)
 173 (total)
 	
 References
 	
 9   (number)
	X Designed Settings (exhibits)
	 X Programs/Events/Activities
	   X schools (e.g., field trips)
	   □ children and teens (non-school)
	   □ adults/general community
	   □ professionals
	 □ Cross-Cutting
	   □ Mass media
	   □ Games & Learning technologies
	 □ Other _____
		X ISI
		   □ aquariums and zoos
		   □ arboretums/botanical gardens
		   X museums and science centers (science-rich)
		   X children's museums
		   □ nature centers (incl. interpretive centers of parks, preserves, etc.)
		 □ General (non-science-rich) museums
		 □ Community/Youth Organizations
		   □ science-rich
		   □ non-science-rich
		 □ Other _____
			□ Internal
			 X External
			 □ Unsure
			
			 Type of Methods
			
			 X Qualitative
			 X Quantitative

Method Code	Method (see key below)	Timing					Sample Size	Comparison or Control Group	Random assignment (if control group)	Instruments included in report?	Notes
		Pre	Retro Pre	During	Post	Delayed Post					
5	Tracking and timing			X			113	No	No	Yes	
3	Interview with adults				X		107	No	No	Yes	
3	Interview with children				X		89	No	No	Yes	
3	Vietnamese interviews			X	X		8 families + 1 leader	No	No	Yes	
6	Observations (other)			X			8 families	No	No	Yes	
3	Mexican interviews				X		16 adults, 14 children	No	No	Yes	
11	Family Science Night case study			X			1	No	No	No	
3	Traveling exhibits interviews				X		10	No	No	Yes	

Methods Key: 1-Survey (self), 2-Survey (other), 3-Interview, 4-Focus Group, 5-Observation (tracking & timing), 6-Observation (other), 7-Video and Audio, 8-Test, 9-Mapping Exercise, 10-Performance Assessment, 11-Other.

19

Example of Framework Worksheet: Summary of Review

Report Citation: Allen, S. (2007). *Secrets of Circles summative evaluation report.* [Report prepared for Children's Discovery Museum of San Jose.] San Mateo, CA: Allen & Associates. Retrieved from http://informalscience.org/evaluation/ic-000-000-007-398/Secrets_of_Circles_Summative_Evaluation_Report	
Total # of Pages	173
Name of Evaluator	Sue Allen, Ph.D., Allen & Associates
Type of Evaluator	External
Name of Institution	Children's Discovery Museum of San Jose (CDM) Evaluation also included interviews with staff from Strong National Museum of Play and Ecotarium.
Type of Institution	ISI—children's museum ISI—(science-rich) museums and science centers
Reviewer Initials	[SK Partners]

The Intervention	
Name of Intervention	Secrets of Circles
Type of Intervention	Designed settings—unfacilitated Programs—schools
Description of the Intervention	The intervention included a 2,600 square foot permanent exhibition at the Children's Discovery Museum of San Jose and two 2,000 square foot traveling exhibitions. In addition, the project team "created a range of public and school programs" (p. 8), such as Family Science Nights.

The intervention itself used interactive exhibits and programs to encourage participants to "explore the basic properties of circles and circular forms (such as curves, spheres, and toroids), to appreciate the elegant use of circles in nature and by people around the world, and to investigate engineering applications of circular shapes such as wheels and gears" (p. 7).

Target audience: "The main audience for the exhibition was children aged 3-10 and their care-givers (parents or other adults). The team designed the exhibition and programs to be particularly inviting to families of Latino and Vietnamese descent, who comprise the largest ethnic groups in San Jose" (p. 7). |

The Evaluation	
Overview	The summative evaluation was a collection of seven studies with an overarching goal to "determine the impact of the exhibition and related materials, in terms of attracting children and their care-givers, engaging them in math/science activities, communicating the special role of circles in science and culture, and welcoming families of Mexican and Vietnamese descent. The team was particularly interested in highlighting any significant differences, between younger children (3-5) and older children (6-10), between adults and children, and between males and females" (p. 8).
Outcomes	• Engagement—General (exhibition use: how thoroughly was the exhibition used, what were the elements that were successful, average time spent at various elements; visitors were "engaged for extended periods with circles-related phenomena and materials: observing, creating, refining, and using circular shapes in collaborative activity" (p.4)) • Cognitive—Content, Skills (understanding the circles theme, understanding the geometric and engineering properties of circles, etc.) • Intrapersonal (personal relevance – visitors connecting circles to their own lives) • Interpersonal (using exhibits in groups, family and group interactions, community participation, adult-child conversations)
Methods	Qualitative, Quantitative
Design	Non-experimental and case study
Comparisons	Comparisons to other projects and studies Comparisons to benchmarks/standards
Findings/Claims	The highlighted findings addressed three main questions: 1. "How did visitors use the *Secrets of Circles* exhibition?" (p. 2) The report compared *Secrets of Circles* usage to Serrell's model (1998) and found that visitors used a high percentage of the exhibit elements and spent a relatively long time in the gallery actively engaged with the exhibits. Compared to Serrell's averages, they used the exhibition "quite thoroughly" (p. 1). 2. "Were some elements particularly successful?" (p. 3) In general, the most attractive elements were interactive ("Ripple Table, Inventing the Wheel, Compass Table, Spin Pictures, and Round and Round" (p. 3)). A bench and three interactive exhibits had the longest holding power ("Circle Videos, Spin Pictures, and Compass Table" (p. 3)). Two of the elements seemed to engage return visitors longer than first-time visitors (Circle Videos and the round boat), and many visitors were able to name personal favorites (p. 3). 3. "Did visitors learn anything from the mathematical theme of the exhibition?" (p. 4) The report specifically mentions that pre- and post-tests were not within the scope of the study, but it suggests the following conclusions based on evidence collected:

	"Visitors of all ages engaged deeply with the circles-related exhibits." (p. 4) "Most adults understood the Circles-related theme of the exhibition." (p. 4) "Children could identify circles in the exhibits, and many could talk about their geometrical and engineering properties in ways that were consonant with the exhibition." (p. 5) "Most adults could identify a physical exhibit with a functional circle in it, though they were not as facile at explaining the relationship between form and function." (p. 5) "Visitors of all ages made personal connections to Circles-related objects and ideas." (p. 6) "Some of the exhibit-related programs engaged children and teachers in thoughtful and intrinsically motivated mathematical explorations and discussions." (p. 6)

Reviewing the Report	
Intervention Rationale	
What Worked Well	**What Could Have Been Done Differently**
Description of intervention: The exhibition is well described with detailed descriptions and supporting photos and diagrams. The description paints a solid picture of what participants see when they enter the exhibit and what they do when they visit. In addition, the intervention seems appropriate for the intended audiences (which are well defined in the report). Rich description of the exhibition (as well as other efforts such as Family Science Night at school), including history, context, staff/teachers involved. References to the literature, mention of prototyping, and collaboration with UCSC support the impression that the intervention was well designed and implemented.	*Maturity of intervention:* Although the exhibitions and related programs appear to be relatively stable, a clear description of how long they have been running would be helpful for judging the maturity of the intervention and its readiness for summative evaluation. *Explicating need and reasoning behind the intervention:* The report implies a need for outreach to the Vietnamese and Latino communities as well as a need to encourage children to engage with and learn about geometrical concepts, but these needs are not explicitly outlined or supported with evidence. Support for these needs probably exists in other documentation (such as the project proposal); a needs analysis may have been done elsewhere, but it would be useful to reiterate it here as part of a complete picture of the overall intervention. That is, help the reader see how this intervention was designed to address existing needs or goals. A logic model, theory of action, or other similar outlining of project goals and assumptions could also have served this purpose.

	Why these evaluation questions: Related to understanding the logic and assumptions underlying the intervention, it would have been useful to have more information about why these particular evaluation questions were selected.

Methodological Rigor and Appropriateness	
What Worked Well	**What Could Have Been Done Differently**
Chain of logic: It is possible to follow the chain of logic within the evaluation study from the intervention goals to the explicitly outlined evaluation questions to the methods and through to the findings.	*Chain of logic:* Connecting the chain of logic through multiple evaluation questions and studies requires some work from the reader. Making these links more explicit though a table or diagram would be helpful.
Transparency of methods/instruments: The author describes the methods in detail and includes instruments in the appendices. In addition, constraints and limitations are discussed in terms of the methods and scope of the project.	*Study design decisions:* The overall designs were described, but not explicitly named. Clearly identifying the study designs and discussing why these designs and methods were chosen over alternatives would provide readers with additional insights into considerations and constraints. This type of discussion would allow readers to evaluate whether those designs were the best available choices for addressing the research questions posed. For example, although the post-test only and pre-experimental and case study designs seemed appropriate for the descriptive questions, using a design that included a control/comparison group and/or pre-post measures would allow for stronger claims relating to questions about visitor learning. The evaluator acknowledges that pre-post was desirable but beyond the scope of the study; more details would have been useful.
Comparisons: The author uses references to the literature to establish definitions and to compare study results to averages and standards in similar settings using similar methodology. These include Serrell's model and Monterey Bay Aquarium's Sharks exhibition (Allen, 2007, pp. 1,10, 16, 17, 20, 21,22; Serrell, 1998; Yalowitz, & Ferguson, 2006).	
Analysis and results: When presenting the analysis and results, the author names statistical tests and explains significance levels used. Results are presented in both narrative and graphic formats as well as supported by data (quantitative and qualitative). Null and negative results are reported, in addition to positive results. Thus, although the author provides her interpretations of findings, readers are also supported in drawing their own conclusions. Interpretations are worded carefully to avoid making unwarranted claims.	*Embedded assessments:* Resource limitations might not render this feasible, but with all the interactive elements in the exhibition, there might have been opportunities to develop some creative measures of learning. Embedded/stealth assessments or transfer tasks/exhibits could have unobtrusively captured data related to visitor learning while they interacted with various elements of the

23

Sensitivity to informal setting: Visitors were unobtrusively observed as they explored the exhibition and most interviews were conducted after visitors had exited the exhibition. In addition, evaluators took steps to accommodate families with children by taking such actions as conducting interviews while walking with families as they continued their visit, and providing toys for children while the adults were talking. When talking to children, a different interview protocol was used to make the experience age-appropriate. Studies that were more intrusive (e.g., involved accompanying and interviewing visitors during their visit) provided a rationale for interrupting the natural flow of the visit in this manner.	exhibition.

Evaluation Uses	
What Worked Well	**What Could Have Been Done Differently**
Quality and transparency of the writing: Although the report is quite long at 173 pages (including appendices), the high quality of the writing increases its digestibility. The report features clear section headings, a logical flow of ideas, appropriate visualizations, and helpful summaries woven throughout. In addition, candid discussions of limitations, scope, negative findings, and unexpected results add a layer of transparency and credibility. *Clear purpose linked to evaluation questions:* The expressed purpose of this summative evaluation is to "determine the impact of the exhibition and related materials, in terms of attracting children and their care-givers, engaging them in math/science activities, communicating the special role of circles in science and culture, and welcoming families of Mexican and Vietnamese descent" (p. 8). In addition, at the beginning of each sub-study, a purpose statement is provided followed by specific evaluation questions. *Identifying indicators of success and comparing them to others:* By being explicit about intended intervention impacts and the	*Identifying evaluation report stakeholders and their needs:* Outlining stakeholders for the report, as well as stating how the report will address their needs, could increase the potential usefulness of the report by allowing readers to see up front how they might use the report. *Providing an expert opinion:* To the extent that the evaluator would feel comfortable and justified in making a judgment call about the overall value of the intervention, this type of summative statement or expert opinion can be informative for stakeholders' decision-making. Along similar lines, it can be helpful to circle back throughout the report to the original need/impetus for the intervention: "Given this evidence collected, how well does the intervention address a compelling need?" *Actual uses of the evaluation:* As with most summative evaluation reports, there is no indication of whether and how this report was used by stakeholders. If possible, it would have been interesting to include plans (if any) for following up on how stakeholders used evaluation information and recommendations.

purpose of the evaluation study, the author is able to identify some indicators of program success by asking questions that address whether, and how well, the intervention achieved the desired impacts. In fact, the evaluator makes a comment in the recommendations section about how the exhibition and programs demonstrated "success in many dimensions" and speculates on factors that may have contributed to these successes (p. 121). Some results (but not all) indicate success by being favorably compared to average results in the field or results from other similar exhibits; these comparisons provide context for evaluating success. *Specific, actionable recommendations:* The evaluator makes several recommendations for addressing areas such as safety, functionality, labeling, marketing, maintenance, storage, outreach, and audience development in order to support existing exhibits/projects and inform future projects (pp. 121-124). Although acknowledging that some recommendations may be challenging and "take years to accomplish," many of the suggestions are specific, actionable, and grounded in evidence from the study (p. 124). *Potential usefulness to several stakeholder groups:* Findings and recommendations could be potentially useful to several different stakeholders including exhibit designers, project teams, funders, and the broader field of informal science education.	