Approaches to Evaluating Faculty Outreach, Part IV:

Demonstration Project—Girls at the Museum Exploring Science, GAMES

Sandra Laursen, Ethnography & Evaluation Research, and Sarah Snow, Museum Studies,
with Cathy Regan, CU Museum of Natural History

1 Introduction

1.1 Overview of the Evaluation Study

Our research unit (E&ER) has carried out a study of the evaluation needs, opportunities, and interests of faculty who have been awarded Faculty Outreach awards from the University of Colorado Boulder Office of University Outreach (OUO). At $5000-8000 per year, these awards support faculty to share their scholarly, creative and teaching expertise with varied local, state or national public audiences. Some projects develop into sizable and sustained efforts over time that may provide high visibility to CU programs, yield scholarly products for faculty, enhance learning for CU students, and attract external funding. Because the aggregate university investment is sizable, the OUO seeks to document the value of this investment to the community and state. They also wish to encourage faculty to think in an evidence-based way about their outreach work, so as to optimize its value to external audiences in ways commensurate with faculty's own needs, capacities and values.

Our study took a two-pronged approach, including:

- interviews with faculty grantees around campus to explore their interest in evaluation, and the evaluation needs and opportunities offered by their projects
- three "demonstration projects" to evaluate Faculty Outreach projects that would provide practical examples and bring evaluation-related concerns, challenges, and possibilities to the surface.

For the demonstration projects, we selected multi-year projects that were well established, that offered evidence of prior success, and whose leaders who were willing to work with us. We attended to variety across disciplines and outreach approaches while seeking examples of approaches common across the university, such as youth experiences, public performances, and K-12 teacher professional development. Here we report on one of these demonstration projects.

While participating in outreach projects clearly offers benefits to CU students and faculty, our work centers on the often greater challenge of documenting benefit to the external audiences targeted by outreach. A separate study addresses the professional outcomes for faculty of conducting outreach, especially its impact on professional advancement (Howe, Davidson & Nash, 2013). For reports on our broader study, see Laursen and Archie (2012), and on the other demonstration projects, see Laursen and Arreola-Pena (2012) and Laursen (2012).

1.2 Overview of the Outreach Project

Girls At the Museum Exploring Science (GAMES) is an afterschool program that seeks to encourage interest in science among preadolescent girls. Groups of 15 girls from fourth and fifth grades at a single elementary school are bussed to the university's Museum of Natural History (CU Museum) weekly for seven weeks, where they explore archaeology, botany, entomology, paleontology, and zoology through hands-on activities using real museum specimens and direct

interaction with scientists and museum professionals. Running in multiple seven-week cycles per year, the program has reached about 350 girls since its inception in 2003-04, targeting local schools that serve high numbers of children from low-income families. The program has been recognized by community grants and by an award for Excellence in Programming from the Mountain-Plains Museum Association.

Sarah Snow (2013) reviewed the literature on afterschool programming and described key features of GAMES that align with research-based best practices. GAMES leaders seek to:

- Promote science through the use of inquiry-based, hands-on learning; presenting science as both a body of knowledge and a process for developing knowledge; and drawing attention to science careers
- Make good use of after-school learning time to spark interest, help close the achievement gap between white and minority children, and help children perceive themselves as capable of doing science
- Support girls and minority students by providing science role models, encouraging girls to express themselves in single-sex environments, addressing stereotypes, and providing access to science and the museum setting to children who might not otherwise have it
- Engage pre-teen audiences to preempt declines in attitudes about science that are common among adolescents, and to inspire them to hold high academic and career aspirations that may include science
- Foster meaningful collaboration through peer-to-peer collaborative learning and scientific mentorship.

Our evaluation questions were developed by E&ER researchers in collaboration with Cathy Regan, a museum educator and the founder of GAMES, and Sarah Snow, a graduate student in museum studies who had worked as a GAMES facilitator and who studied GAMES outcomes for girls as her masters thesis. Our questions about GAMES were:

- What outcomes for girls of participation in GAMES are evident in historical records of GAMES data and activities?
- What lasting outcomes for past participants can be detected?
- What advice can be offered to improve future opportunities for studying the short- and longer-term outcomes of GAMES for girls who participate?

A second set of evaluation questions grew out of the broader project to examine the needs, interests and opportunities for more robust evaluation of OUO-funded Faculty Outreach projects.

- What methods may be used to probe these questions, and what considerations enter into selection of methods for a particular study?
- How do the selected methods work in practice to gather information, with what results, what investment of resources and what potential for sustained independent use in the sponsoring programs?
- What can be learned from the pilot project that is useful in setting expectations and implementing evaluation appropriate for funded Faculty Outreach projects?

These questions were developed in collaboration with the outreach project leaders and OUO staff, and with awareness of recent scholarly work on university outreach and engagement (e.g., Fitzgerald, Burack & Seifer, 2010).

2 Study Methods

We discussed a wide range of ideas before settling on a study plan. As leaders, Regan and Snow viewed the program as well-honed and did not see great need for formative evaluation. From girls' verbal comments and responses on end-of-program surveys, they had a sense that girls learned some new science concepts and left the program with positive feelings and interest in science, at least for the short term. They were most interested in the possible lasting effects of GAMES on girls as they progressed in their academic careers: increased interest in science courses and careers; confidence, persistence and success in their science courses; and self-concept as a student who is good at science.

Several conceptual features constrained the study design. The young age of participants (grades 4-5) and alumnae (still school age) requires both age-appropriate study methods and special care to protect children's rights as human subjects. GAMES participants attended multiple elementary schools; cohorts may not remain together when they move to middle school. Since the program is run through schools, the school district's permission and assistance were required. GAMES did not have participant data that would enable us to identify and contact alumnae individually.

2.1 Short-term outcomes

To consider students' short-term outcomes such as growth in knowledge, interest, and confidence, we reviewed information already gathered from GAMES participants and considered ways to build on this information for more systematic data collection. For example, in a common opening-day activity, students drew and discussed their image of a scientist; while girls kept the notebook where they made their drawing, notes were kept about the ideas girls shared. At the end of the program, girls completed a short survey about their attitudes, confidence and content knowledge. We discussed ways to better document such outcomes in a systematic yet nonintrusive manner, so that the experience did not disrupt programming nor feel like a school test for girls. We focused on the potential of "embedded assessments" such as the Draw-a-Scientist activity, often used to examine children's stereotypes of scientists (Chambers, 1983).

The Draw-a-Scientist activity and immediate post-program surveys had been routinely gathered for formative feedback and used to make mid-program adjustments. These data had not been systematically analyzed across cohorts, so rather than collecting new data for a single school year, we elected to review the accumulated body of data as historical evidence of immediate outcomes for girls. Sarah Snow analyzed these sources, focusing on 93 questionnaires from 2005-2012. She tallied multiple-choice responses and conducted thematic qualitative analysis of girls' short-answer responses. Selected results from her analysis are reported in Section 3.

2.2 Longer-term outcomes

We also considered ways to gather data about more lasting influences on girls' interest, confidence, science identities, and academic choices, focusing on the possibilities in a retrospective design that would examine the current status of past GAMES participants. Our options here were limited by a lack of records of the names and addresses of past participants. The museum did not keep records of GAMES participants' names and contact information—in part because the selection of girls and collection of permission slips was done at school, through

the chaperoning teacher, and in part to protect students' confidentiality (e.g. for students who might be undocumented). We contacted schools to see if they retained lists or permission slips that would let us identify participants. They indicated that retention of this information was not routine nor likely to go back far enough in time to be useful for a retrospective study. Thus our ability to single out GAMES participants from the much larger set of students the same age was limited to girls' own memory and willingness to self-identify.

Because all the girls came from Boulder Valley School District (BVSD) elementary schools, we considered the feasibility of examining formal academic records of GAMES alumnae, to look for evidence of students' persistence and success in science classes in comparison to non-participant peers. District officials indicated that, while they were interested in GAMES outcomes, they would require a strong rationale that this study approach was likely to be fruitful and not, in their words, a "fishing expedition." This method was also ruled out by the lack of participant lists.

Ultimately, we chose to carry out focus groups with GAMES alumnae of middle and high school age, those who were 2-8 years past their GAMES experience. This would enable the interviewer to probe girls' memories of GAMES and any changes in their attitudes and perceptions about science that they might attribute to GAMES.

In order to set up focus groups, however, we had to reach girls through the school district. To take advantage of this contact for additional data, Snow developed a short survey that was mailed by BVSD to all girls in grades 6-12 who had attended the elementary schools where GAMES was offered when they were in grade 4 or 5, and who were thus possible alumna of GAMES. The survey asked girls about their participation in GAMES and other science programs, and about their attitudes, confidence, and career aspirations. They were asked to indicate their willingness to participate in a focus group. The survey was accompanied by an extensive set of youth assent forms and parent consent forms, in Spanish and English as required by the school district. Seeking to minimize study fatigue among its students, BVSD did not give permission to survey non-participants, but some did in fact reply to the survey. Both the survey responses and a set of focus groups were used as data sources for the retrospective study. A total of 43 students (30 from GAMES, 13 others) completed the survey and eight students participated in four focus groups. Snow carried out descriptive quantitative analysis of the survey and thematic coding of the transcribed focus group data. Her results for these analyses are included in Section 3.

2.3 Adult perspectives

In addition to the student retrospective data, we pursued interviews with chaperones from each elementary school. These teachers, staff or parents helped to organize the program and rode the bus with the girls to the museum. Many had participated for several years and could provide a perspective based on their observations of the girls during and after the program. Five of 12 chaperones contacted participated in this aspect of the study. The interviews were transcribed and coded; key results are reported in Section 3 along with results of student studies.

We discussed but did not carry out a study to examine the impact of participation on the CU students and scientists who host GAMES schoolchildren in their labs or museum spaces. Such a study could be carried out in the future and based on prior work on scientists' involvement in outreach (see e.g. Thiry, Laursen & Hunter, 2008; Laursen, Thiry & Liston, 2012). Family responses, such as parents' observations about their daughter's learning and their own views of science as an education or career goal, offer another possible study approach that might be incorporated into the museum's efforts to involve families more extensively.

3 Evidence of Desired Outcomes

We initially developed our design considering very broad research questions about the short- and long-term outcomes of GAMES participation for girls, framed by our targeted evidence sources: previously collected post-program questionnaires and a new, retrospective study using student surveys and focus groups as well as teacher interviews. However, after analyzing the data, we found it made more sense to organize the results by major types of outcomes and to draw upon all available data sources in answering each of these more focused questions:

(1) Do GAMES participants believe themselves capable of participating in and contributing to scientific work?

(2) Do they maintain a positive attitude towards science? Do they like science as much as (or more) than they did when they participated in GAMES?

(3) Do they continue their academic coursework in science and/or pursue science opportunities outside of the classroom?

In addition to assessing the quality and nature of evidence about each type of outcome, we also sought to attribute these outcomes to GAMES. Without a link to the program itself, it would be difficult to claim that GAMES was the cause of any lasting impact on participants. However, the presence of positive attitudes towards science along with a pattern of attribution to GAMES would indicate that the program was meeting its long-term goals. Likewise, attribution of negative attitudes would suggest that the program was failing to meet its goals, or even doing harm.

3.1 Confidence

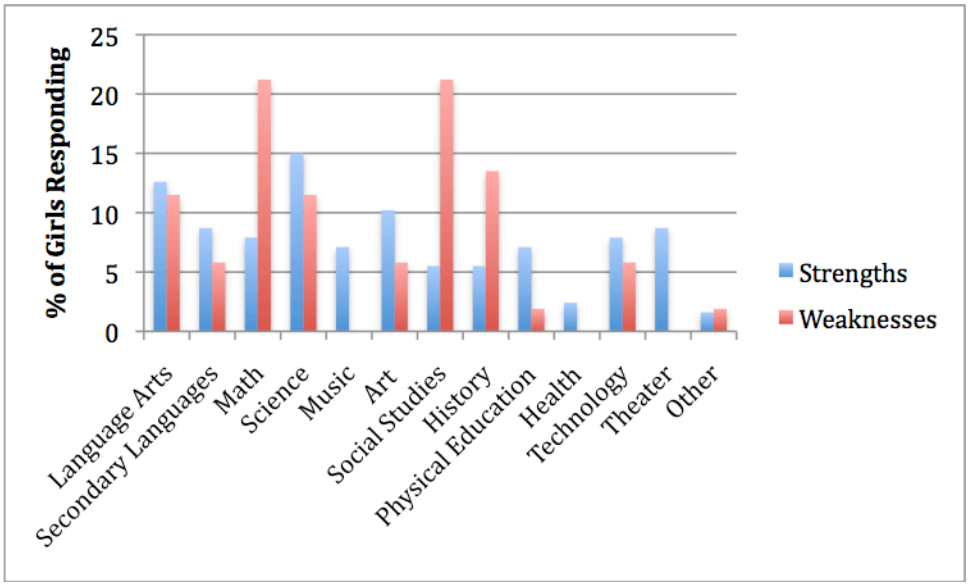
Confidence is an important indicator of persistence in science. Assessment of confidence addressed not general self-esteem, but science confidence specifically: Do GAMES participants believe themselves capable of participating in and contributing to scientific work?

Assessment of girls' growth in confidence came from two main data sources, both from the retrospective study. First, Snow offered a global assessment of girls' overall level of confidence for each participant in the focus groups, based on discussions of their own abilities in science (past and present), any concerns about upcoming courses, and their future goals for college degrees, careers, and so on. Of the eight girls who participated, five demonstrated high confidence, two medium, and one lower confidence. One interviewed teacher noted that GAMES participants showed more confidence in class than did non-participants. Another teacher offered this comment:

> Overall the girls that have attended this program have felt empowered. They are much more vocal in science class. This last Wednesday, our fifth graders had an opportunity to dissect a lung. The GAMES girls were the first to volunteer and begin the dissecting.

Second, the follow-up survey probed girls' confidence by asking what school subjects they were "good at." More than two thirds of GAMES alumnae reported that they were good at science (Figure 3.1). In fact, science was the most commonly reported strength among this group with other favorite subjects, language arts and art falling close behind. This indicates that past GAMES participants feel confident in their abilities to perform science.

Figure 3.1: Strength and Weakness in School Subjects of GAMES Alumnae



3.2 Attitudes toward Science

Examination of girls' attitudes assessed both their general attitude about science and self-reported changes over time since they participated in GAMES. Data sources include the immediate post-survey and both follow-up sources, the survey and focus groups. Table 3.1 presents attitudes as determined from these measures. The table is structured to facilitate comparison of values *within* one study type; comparisons of results *across* study types are not meaningful because of the small and unrepresentative samples that could be gathered.

Table 3.1: Comparison of Change in Science Attitudes from Multiple Data Sources

Compared to before GAMES, I like science…	Immediate post-survey (n=92)		Follow-up survey of alumnae (n=30)		Focus groups with alumnae (n=8)	
	# of girls	% of girls	# of girls	% of girls	# of girls	% of girls
More	78	84%	14	47%	4	50%
The same	12	13%	10	33%	1	13%
Less	1	1%	5	17%	1	13%
Left blank, unclear	2	2%	1	3%	2	25%

Most participants report positive and/or improving attitudes about science, both in the short and longer term; the fraction of students reporting declines in their attitude about science is small in all samples. It is interesting, but perhaps coincidental, that the proportion of students who reported increases in their liking for science at the time of GAMES (84%) is essentially the same as those who later reported liking science the same or more (80%). While the immediate and follow-up samples cannot be matched and may not represent the overall population of girls, we can conclude that at least a subset of past GAMES participants remain keen on science and their positive attitudes are sustained or continue to develop.

In corroboration with the girls' self-reports, two teachers noted that GAMES participants were more interested in science and excited about learning than non-participants.

> Because they had taken a lot of those skills with them and they had already had a lot of experience, they could easily apply that to what we were doing in class. …They had a lot of things to connect with, which was really great.

Another teacher reported that they showed more confidence in class, while two others reported some envy among students who did not have the opportunity to participate in the program. This suggests that participants' positive conversation or effects of the program were perceptible to others outside of the experience. One chaperone shared by e-mail: "The girls have come back to school very eager to share their newly found knowledge with their classmates. One student, I recall, stated that she had no idea that women could be such great scientists!"

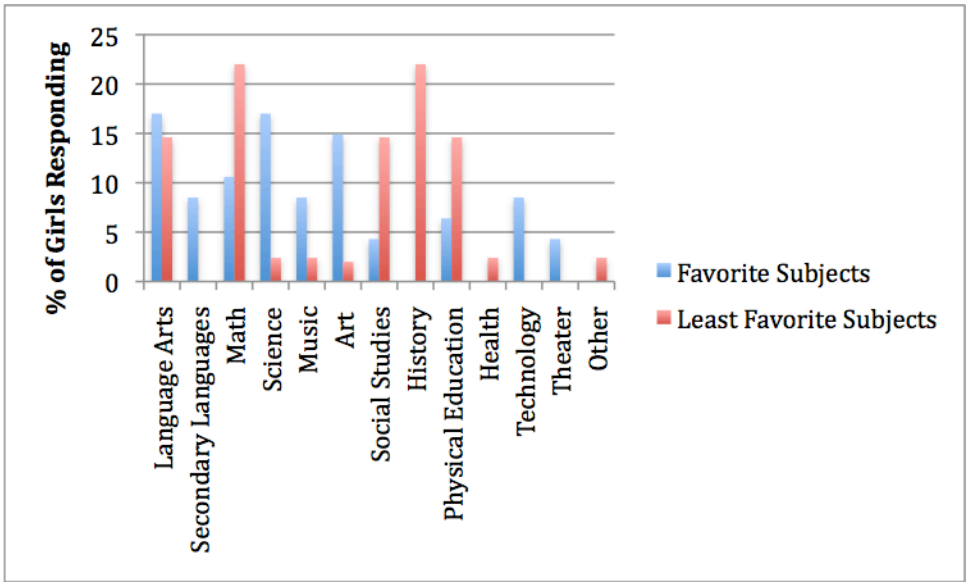
3.3 Academic and career aspirations

Both interest and confidence are personal factors that may influence girls' aspirations. The study also directly probed girls' interest in science as an academic and career interest: do they continue their coursework in science and/or pursue science opportunities outside the classroom?

In general, GAMES participants had not participated in high rates in other out-of-school science programs (41% had, 48% had not, 10% did not know). Given GAMES' target population, this may be because they had not had the opportunity. Comments by focus group participants suggested that some opportunities increased in high school, such as after-school clubs, but others, such as Science Olympiad, were more available in middle school. It is not clear if students considered these school-related programs in their answers.

On the follow-up survey, girls were asked to indicate their current favorite and least favorite subjects in school (Figure 3.2).

Figure 3.2: Favorite and Least Favorite School Subjects of GAMES Alumnae



Common favorites were language arts, science and art. This indicates that girls' positive attitude toward science is not restricted to out-of-school settings, but also applies to science as a school subject. Interestingly, math was the leading least-favorite subject, suggesting room for

improvement in this key STEM arena, perhaps through connection to science. Comparison of Figures 3.1 and 3.2 shows that, in general, students seemed to prefer subjects that they perceive as their strengths.

On the follow-up survey, girls generally expressed high educational aspirations. Of the 30 respondents, 72% planned to graduate from high school (plus 14% planning to earn a GED), 90% planned to go to college, and 45% to gain education beyond college. The fact that college aspirations exceeded high school graduation may reflect some naïveté about college entrance requirements, or may simply mean that students did not check all applicable boxes, taking high school graduation as a given if they had already checked college as an aspiration.

Information about girls' career interests was gathered on the immediate post-program surveys and on the follow-up survey, and in the focus groups. All three data sources suggest that girls had a high level of interest in science careers, both at the time of participation and later on. It is unclear to what degree this can be attributed to GAMES participation, to selection of science-interested students for the program, or to sampling bias toward science-interested students.

Table 3.2 compares results on career aspirations from the three sources. We caution strongly that quantitative differences should not be interpreted as meaningful, as the samples are small for both alumnae data sources and not statistically representative. Focus group and interview data provides a sense of the weight of opinion but cannot be treated as a statistical measurement.

Table 3.2: Comparison of Career Interests from Multiple Data Sources

Career aspiration	Immediate post-survey (n=92)		Follow-up survey of alumnae (n=30)		Focus groups with alumnae (n=8)	
	# of girls	% of girls[1]	# of girls	% of girls[1]	# of girls	% of girls
Science	62	77%	15	50%	5	63%
Non-science	19	24%	15	50%	1	13%
Unsure or blank	10	12%	4	13%	2	25%

[1]Percentages add to more than 100% because some girls noted multiple career goals.

In the immediate post-surveys, the five focus disciplines appeared commonly in students' career aspirations (e.g. botanist, zoologist). A question asking students to identify these disciplines had appeared on the same page, thus it is possible that girls' ideas were prompted or constrained by the earlier question. Other science careers were veterinarian and doctor. Non-science aspirations included teacher, entertainer (singer, actress), police officer, artist, professional baseball player, and immigration lawyer.

Career interests stated in the follow-up surveys included doctor, biologist, psychologist, veterinarian, and engineer. Non-science careers were similar to those cited earlier: artist, teacher, entertainer, media professions, and lawyer.

The sample sizes are too small to draw conclusions about trends over time in girls' career interests. However, the high level of science interest immediately following the program is one indicator of program success. It would not be surprising if interest declined somewhat over time as girls are exposed to other fields and develop new interests. However, the longitudinal samples are quite small and we cannot determine whether they are representative of the population.

Among notable changes witnessed in their students, teachers reported more GAMES-related science fair projects, as well as GAMES-inspired career aspirations. "They [were] seeing that people do this for their job," noted one teacher. "It's very exciting and very, very important." Another teacher described a student who had gone on to study engineering at CU and returned to the elementary school to talk to current students about GAMES and her own career choices. Still another teacher remarked of a student:

> She just went to town on those bugs and began her own bug collection. And her mom's like, "Why'd you take her to that? Now I've got bugs all over the house," so I'm like "Yay!" …I heard transformation in the way she talked about future goals—moving from the waitress job, which—there's nothing wrong with waitressing—but that she all of a sudden was thinking, 'Hey, you know it would be really cool if I could study all these little parts of bugs.'

3.4 Other Outcomes

The post-program surveys offer evidence of some science content learning. Girls were asked to match each of the five scientific disciplines studied in GAMES (central science areas of the CU Museum) with a word describing its focus: archeology to ancient humans, botany to plants, entomology to insects, paleontology to fossils, and zoology to animals. A total of 56% of girls correctly identified all five sciences. Zoology was the most readily recognized, with 94% of respondents correctly matching this science, and the percentage correct on each individual science ranged from 67%-74% for the other four sciences.

From teachers' perspective, participants gained broader perspectives on science, the different types of science, and the people who participate in science. Especially important to these new perspectives was the presence of female role models—mentioned in all five interviews.

> There's not a lot of exposure to females that are in high levels of professional careers that they're able to look upon and say, 'Oh yeah—I could do that' or 'Oh yeah, that's something that I really like to do,' and to see somebody who's modeling that for them. So that was one of the really big, big eye-openers for a lot of the girls… they assumed that when you say that you're a doctor, that that meant that you're a boy and that you were a man.

Other common gains suggested by teachers included exposure to higher education and the university campus, increased feelings of empowerment and confidence, and the awareness of diversity among potential careers. These observations are consistent with GAMES' conceptual framework and suggest that girls receive opportunities to relate personally to science and to come to see themselves as capable of contributing to science.

These outcomes, however, are not independent of teachers' pre-selection of students to participate. They reported a variety of practical and conceptual approaches to student selection: all five said they looked for some form of interest in science, and four tried to seek out students who did not have similar opportunities: those who had not visited a museum or college campus, those from economically disadvantaged backgrounds, less supportive family situations, and diverse cultural perspectives. This suggests that the opportunities are reaching students who can benefit, but also who have pre-existing interests. Thus it is more difficult to attribute outcomes to GAMES alone.

3.5 Attribution of Outcomes to Participation in GAMES

The outcomes noted above—confidence, positive attitudes about science, and career interests—are complex, formed over time under many different sources of influence: the family, school, peers, and outside activities, as well as cultural values and media exposure. Participation in GAMES or any out-of-school-time science program is only one possible factor. Thus it was important to attempt to connect these outcomes with GAMES participation. With small sample sizes, and lacking a comparison group, it is not possible to make statistical correlations that might suggest causality. Instead, the study directly explored this attribution by looking for evidence that girls linked their statements about outcomes to their participation in GAMES. The follow-up questionnaire and focus groups both probed girls' memories of GAMES as a means to determine whether these memories were strong enough, and positive enough, to contribute to their current science attitudes and perceptions.

On the post-program survey, girls described what they liked most and liked least about the program. The middle column of Table 3.3 shows the number of distinct aspects noted in student responses at the time of the program (each response may contain multiple ideas).

Table 3.3: Aspect of GAMES Girls Liked Best, Post-program, Compared with Memories of GAMES, Follow-up

Aspect liked most	Post-program survey (N=92)	Follow-up survey (N=30)
	% of all responses (n=132)	% of all responses (n=171)
Hands-on activities	20%	15%
Science in general, a specific discipline	17%	28%
Museum objects or collections	17%	16%
Learning or studying, new experiences	16%	16%
Visiting or talking with scientists	11%	9%
Everything	9%	6%
Fun	3%	2%
Visiting campus	3%	2%
Miscellaneous/other	3%	5%

Most-liked aspects included hands-on activities (including "using tools" and "touching things"), science in general or particular disciplines, learning or studying, and references to museum objects and collections. A number of girls reported liking "everything."

Girls also wrote in the aspect they liked least about GAMES (Table 3.4). The most common answer was "nothing," meaning they liked everything. Because sessions are organized around particular science topics, the disciplines are closely associated with both likes and dislikes. Mental discomforts included "to have to stick a needle through a grasshopper," "the smell of the rooms – yuck," and "seeing snakes."

Table 3.4: Aspect of GAMES Girls Liked Least, Post-program (N=30 girls)

Aspect liked least	% of all responses (N=201)
Nothing	18%
A specific discipline	42%
Non-hands-on activities (writing, talking, waiting)	12%
Mental discomfort (things that were "gross")	8%
Missed opportunities (due to absence from a session, not enough time)	6%
Physical discomfort (too much walking or standing)	4%
Miscellaneous/other	8%

In the follow-up survey, girls were asked to share something they remembered about GAMES. This was partly a check on whether they were accurately remembering that they had taken part in GAMES, but also a way of relating outcomes to participation. Not surprisingly, these often mirrored aspects mentioned as "most liked" on the immediate post-surveys (see right column of Table 3.3)—but negative memories were also mentioned. They described GAMES program elements that were particularly enjoyable, memorable, and influential on their thinking, especially hands-on activities and the incorporation of authentic objects and experiences.

Other common memories included exploring the museum, spending time on the university campus, and having fun. The match of the elements that girls reported as important immediately following the program, and what they remembered later, shows that they could remember the experience vividly and thus ascribe particular changes to it.

Another line of evidence comes from the focus groups. Again, girls' memories of GAMES were used as a way to verify their participation and document that the experience had been personally significant. They mentioned specific experiences similar to those highlighted in the immediate post-program survey, and recalled the science tools they had received in what one girl called her "schwag bag" to take home. For example, 7th-grader Abby remembered enjoying the plant unit, touching fossils, dissecting dead birds, and perusing a traveling museum exhibit about human consumption. She did not like spending so much time on the bus and had to forgo the entomology unit because of her fear of spiders. Addy, also a 7th-grader, remembered dissecting flowers, using microscopes and pinning bugs. She was particularly impressed by getting to see extinct animals—including a passenger pigeon and a parrot—and the delicious smelling eucalyptus in the botany collection.

Snow also systematically examined the relationship of girls' attitudes to their participation. Treating her eight participants as individual cases, Snow categorized their overall attitudes and views of science, then assessed the nature and strength of the relationship of these attitudes to what girls said about how GAMES had or had not influenced their confidence, attitudes, and interest in science. Her categorization of the cases is shown in Table 3.5.

Table 3.5: Categorization of Case Studies by Attribution of Outcomes to GAMES

	positive attitude about science		negative attitude about science	
	+ + positive attitude positive attribution	+ - positive attitude lack of attribution	- - negative attitude lack of attribution	- + negative attitude positive attribution
Strong relation	Addy (gr 7) Kami (gr 7)	Camille (gr 10)	Abby (gr 7)	
Weak relation	Abby (gr 9) Meghan (gr 9) Julia (gr 10)	Mia (gr 7)		

Snow found no cases in which GAMES participation was negatively related to outcomes—that is, no cases in which GAMES had appeared to cause a loss of interest or positive attitude. The cases are thus categorized by positive attribution or lack of attribution, and by stronger and weaker relationships.

Snow categorized a positive and strong attribution in cases where students seemed to link their participation in GAMES to their enhanced or sustained interest in science. Some did so directly —for example, Addy, a 7th grader, commented, "For a while before that, I wanted to be a zoologist— but after that, I was really more sure that I wanted to be a zoologist." Here she recalled her prior interest in a science career but also connected her participation in GAMES to a strengthened interest in that profession.

Some weaker relationships were also deemed to be positive attributions, but the evidence was less direct. For example, 10th-grader Julia had gotten a lot of use out of her toolkit over the years. She still had the measuring tape, though the other tools had been broken or lost. Although her positive experiences in GAMES did seem to help maintain an academic interest in science, her long-term career goals were not affected by the program. She wished she had been exposed more generally to higher education, suggesting, "It'd be cool to learn more about college and going and living on campus and stuff. You know, meeting professors and figuring out how it works and class at college and what not. That would be kind of cool."

Mia, a 7th-grader, saw herself as a marine biologist and hoped to study at the University of Hawaii. Although she reported still having all of the tools, there was little evidence of a pattern of correlation between Mia's participation in GAMES and her current interests in marine biology. Thus she was categorized as having a positive attitude toward science, but without attribution to her GAMES participation. Details of the other cases can be found in Snow (2013).

Outside observations by teachers provided further evidence of changes in student attitudes, behaviors, and aspirations that teachers saw as directly tied to specific GAMES experiences.

4 Lessons learned

Here we summarize three kinds of learning that emerged from this study:

- Findings about the GAMES outreach project itself (4.1)
- Specific insights into the kinds of evaluation approaches that appealed to project leaders

and that worked in practice (and that did not), and thoughts on how these approaches will continue to inform this project and/or serve as useful models to others (4.2)

- General insights that may inform other outreach projects and strategies of the Office of University Outreach (4.3)

4.1 Findings about the GAMES program

First we comment on what was learned from each type of analysis.

Overall, analysis of immediate program feedback indicates that GAMES has positive effects on participants' attitudes towards science. The majority of girls who complete the program report liking science more than they did before, and they express an interest in pursuing science as a career. Key components of the program's conceptual framework—including hands-on activities and interaction with authentic objects and experiences, and with working scientists—were referenced as contributing factors to the short-term success of GAMES.

Responses to the follow-up survey provided evidence that past GAMES participants have confidence in their scientific abilities and believe themselves capable of participating in science,. They maintain quite positive attitudes towards science—though, understandably, perhaps not as overwhelmingly positive as immediately after the program. These data did not provide evidence that past participants were continuing their academic coursework in science or pursuing other science opportunities outside of school. The comparison group of students who replied to the follow-up survey but were not GAMES alumnae was too small for meaningful comparisons to be drawn. We do not know that the sample is representative: the evidence shows some good outcomes that can be achieved, but not how widespread these outcomes are.

From student focus groups, we document that seven of the eight students interviewed demonstrated (1) some level of confidence in their abilities to perform science, (2) a continuation of academic coursework and/or pursuit of outside opportunities in science, and (3) the maintenance of generally positive attitudes towards science. A few students were able to describe direct links between these outcomes and their experiences in GAMES, while more tenuous ties between program participation and positive outcomes were observed in some other cases. Only three of the eight cases demonstrated no evidence of a linkage—either positive or negative—between GAMES participation and present attitudes, and in no instance did the program appear to have a negative impact. Again, we cannot characterize whether this sample is representative, but we can document that positive impacts on students are possible and can be detected in retrospective reports offered 3-6 years following participation.

Observations by teacher chaperones provided even more patterns of relationship between the GAMES program and desired outcomes. In interviews, these long-time observers of the program offered evidence of changes in student attitudes, behaviors, and aspirations that they saw as directly tied to specific GAMES experiences, drawing upon multiple observations of many students over time. Thus these results generally corroborate student reports. However, the teacher sample was small and may be biased toward firm supporters who were willing to take time to participate in hopes that the program would continue to benefit their students.

Thus, despite the very significant challenges encountered in carrying out this study (see 4.2), positive outcomes were detected. Although these outcomes are not generalizable, in many cases they were corroborated by multiple forms of evidence. The use of multiple methods helped to solidify the findings despite the small sample sizes, lack of matched pre/post samples, or a non-

participant comparison group. In general, the findings suggest that the design of GAMES is effective for accomplishing its goals. It appears that young girls benefit from participating, including in ways that may have impact well beyond the experience itself.

In addition to the summative evaluation findings described, we derived some useful formative feedback from the data and from our conversations. Teachers in particular suggested improvements to the logistics and transportation, and to clarity of selection guidelines. Both teachers and students suggested putting more emphasis on college expectations, college life, and the attainability of higher education. Program leaders are considering incorporating this into follow-up programming for the girls and their families. Other ways to follow up with girls were also discussed, for example engaging GAMES alumnae as "station leaders" for a hands-on activity at a family day. Ideas for a "second tier" program for GAMES girls include new topics outside the museum's specialties, such as chemistry, physics and engineering, perhaps conducted jointly with iSTEM, physics outreach, or Science Discovery. Finally, teachers all expressed a vote of confidence through their hope to see the program expanded to more students and schools.

Sharing these findings with the participating scientists should provide them with insight into how they could increase their impact on the girls' confidence and interest in science. Leaders discussed the potential of a professional development session (e.g., a working breakfast) for the scientist participants, where program leaders might share feedback from the girls and promote sharing of successful strategies for working with girls and reflecting on the benefits to themselves of participating in GAMES.

4.2 Evaluation capacity-building for GAMES

As each of the demonstration projects has revealed, carrying out a thoughtful and ambitious evaluation of an outreach program is a great deal of work. We emphasize that, while the study was conceptualized and developed as a collaborative exercise, neither the Museum team nor the E&ER team alone could have conducted the work represented here.

In many respects, GAMES was a best-case scenario among the three demo projects. Current interest in the outcomes of informal science experiences for youth is high. The topic is timely and interesting to the field, yet not well explored—thus the study had a chance to contribute knowledge new to the field as well as gather useful feedback to the program itself. The E&ER team had another project on out-of-school-time science programming for youth and was familiar with the literature and possible models for study design. The project was mature and led by its founding director, a museum staff member whose professional work focuses on education, and who therefore had professional interest in girls' outcomes. A graduate student already well-versed in the program stood ready to invest time and effort to carry out a thesis-quality study. She remained patient and flexible while rising to the substantial demands of study design and execution. The project built on already strong relationships with the school district, including the participating schools, their teacher chaperones, and the research and assessment office. The district was generous in giving advice, preparing mailings and providing access to participants.

Nonetheless, numerous challenges were encountered in carrying out this study. While some details were specific to our study design, the issues that they raise are quite general:

- The Museum did not keep records of the names and contact information of its past GAMES participants, so the population for follow-up study could not be uniquely identified. This also limited the choice of study designs (e.g. no pre/post matching was possible).

- The retrospective study involved procedures beyond those routinely part of participating in the outreach activity itself, therefore IRB approval was required. Work with children requires parental permission and other extra human subjects protections.
- Access to the girls could be accomplished only through the school district, therefore IRB approval was required from both the district and the university, with sometimes conflicting expectations. The time required was underestimated, and several iterations were required, which together delayed the study's onset.
- The district restricts research in the schools to certain times of the year. Combined with other timeline issues, this delayed in-school data collection until the second year (fall 2012).
- Several aspects of the procedures put in place for human subjects protection likely contributed to low return rates.
    - The mailed recruiting packet of 14 pages (cover letter, questionnaire, consent, return instructions and envelope, Spanish translations) may have deterred or confused potential participants. Several potential respondents returned data but did not include all the required forms of permission, so their data could not be used.
    - The district prepared and sent the unsolicited mailing so that students would remain anonymous to the researcher unless they self-identified to participate. However, the district combined this mailing with that for another study that was behind schedule by three weeks. Thus the survey 'return-by' date was close to the date the mailing arrived. Some potential respondents may have believed it was too late to participate.
    - An extra review process after the initial approval delayed the scheduling of focus groups. Students who had initially elected to participate may have lost interest or made other commitments by the time the researcher was able to contact them.
- The district did not wish to treat non-participants who received the mailing as a study comparison group. The absence of a comparison group limited the extent to which data could be interpreted and generalized.

Enumeration of these issues is not intended to criticize IRB, district or university requirements. Protection of sensitive subjects is a fundamental ethical responsibility in evaluation research, and working with these requirements is an important learning experience for graduate students and other researchers new to this arena. However, these examples highlight the effort and long lead time required to carry out this kind of study, and the ways in which study design and implementation may be constrained. Indeed, our experience in this project provides graphic testimony as to why studies of the long-term effects of out-of-school programs are so rare.

Some learning from this study will have immediate impact on GAMES' formative evaluation practices. Snow (2013) offers several useful suggestions to improve program record-keeping and the use of embedded assessments such as the Draw-a-Scientist activity. These simple methods, which do not require IRB approval, can be optimized for routine monitoring of short-term outcomes. More systematic collection and analysis of the post-program survey can strengthen the conclusions drawn from the immediate post-surveys and begin to build a record for possible comparison with longitudinal or retrospective data. Keeping participation records (e.g. copies of permission slips) and asking permission for the Museum to contact girls or their parents will make it feasible for a researcher to carry out the study without school district involvement. Ideally this can be done without compromising the trust of parents in sharing

information with the Museum. Program leaders will consider the potential benefits of documenting girls' socioeconomic status and racial/ethnic identity to assess whether they are increasing confidence and interest in girls from backgrounds underrepresented in science careers. It is also clear that future studies will benefit if less time elapses between instances of communication between the museum and participants.

The use of multiple methods is a strength of the study. Initially we were most interested in focus groups, but our means of access by mailing all students meant we could administer a survey while also recruiting focus group participants. The two forms of data proved complementary; the survey data pool was larger, while the focus groups fleshed out not only the outcomes but the kinds of trajectories girls may take as their science interests evolve. We learned much about the potential of both methods that can guide future research. Snow's (2013) analysis of the literature on out-of-school-time programming for girls and her placement of GAMES in this framework will be an asset in fund-raising and further developing the program, and we hope that her analysis will serve as the basis of a future peer-reviewed publication describing the program.

4.3 Implications for OUO practice

Our other reports document one challenge inherent in evaluating many outreach programs, in that the value of the data is not always clear to outreach teams. With good reason, outreach teams do not want evaluation processes to interfere with doing the outreach itself, and the value of the data for the program is not necessarily clear to them. In this case, intrinsic interest was not a barrier; data on youth participant outcomes were seen as valuable to the program. As we have noted, evaluation data may be perceived as more useful, and the time spent gathering them more worthwhile, by the university's outreach professionals than by regular faculty. Yet the outreach team viewed outside evaluation expertise as essential to moving forward.

Should OUO implement more rigorous or standardized evaluation expectations, providing in advance a framework for evaluation will be important. But the major implication for OUO practice addresses the time and resources required to carry out meaningful evaluation of outreach programs that involve youth. It is not surprising that most youth outreach programs will not be able to carry out such studies alone. Yet, ironically, these are the outreach programs whose longer-term outcomes we would most like to learn about: such programs have high symbolic and marketing value, and they may provide a pathway for youth and families to come to know the university, see it positively, and pursue higher education. Reaching out to underserved student groups, as GAMES does, may be especially important, offering young people the potential for a truly life-changing experience.

Given these considerations, we suggest that there is significant potential in a joint evaluation of multiple university youth outreach programs—perhaps a subset focused on youth development outcomes, or those in STEM, several of which are supported by OUO and Continuing Education (Science Discovery, GAMES, Math Circles, etc.). Such a project might leverage resources, initiate collaborations, and enable comparative and longitudinal data-gathering that is of high interest to the field. Because they are supported by significant university resources in the form of intellectual, facilities and personnel investments, university-sponsored outreach programs to youth tend to be long-running and positively perceived in the community, and may be good targets for such a study. Initial university investment in exploring interest and possibilities for such a study may make it possible to develop a design that can attract external funding.

5 Acknowledgments

Deane Bowers provided helpful comments on the study design and results. Rebecca Crane assisted with initial data collection. Brigitte Mutter of the Boulder Valley School District and Claire Dunne from the CU IRB office provided valuable assistance.

6 Appendices

Snow (2013) provides the full human subjects protocol, consent and assent forms. We include here the main student instruments.

A) Post-program questionnaire

B) Follow-up survey

C) Student focus group and teacher interview questions

7 References Cited

Chambers, D. W. (1983). Stereotypic images of the scientists: The draw-a-scientist-test. *Science Education*, 67(2), 255-265.

Fitzgerald, H. E., Burack, C., & Seifer, S. D. (eds.) (2010). *Handbook of Engaged Scholarship: Contemporary Landscapes, Future Directions* (vols. 1 & 2). East Lansing: Michigan State University Press.

Howe, K., Davidson, K., & Nash, C. (2013, January). How the Faculty Merit Evaluation System at the University of Colorado Boulder Credits Outreach and Engagement Activities. [Report to the Office of University Outreach] Boulder, CO: University of Colorado Boulder.

Laursen, S. (2012). *Approaches to Evaluating Faculty Outreach, Part III: Demonstration Project—Evaluation of Teacher Professional Development Workshops on Sound and Hearing Health*. [Report to the Office of University Outreach] Boulder, CO: Ethnography & Evaluation Research, University of Colorado Boulder.

Laursen, S., & Archie, T. (2012). *Approaches to Evaluating the Outcomes of Faculty Outreach, Part I: Needs, Opportunities and Challenges.* [Report to the Office of University Outreach] Boulder, CO: Ethnography & Evaluation Research, University of Colorado Boulder.

Laursen, S., & Arreola-Peña, M., with Posnick, P. (2012). *Approaches to Evaluating Faculty Outreach, Part II: Demonstration Project—Evaluation of CU Contemporary Dance Works*. [Report to the Office of University Outreach] Boulder, CO: Ethnography & Evaluation Research, University of Colorado Boulder.

Laursen, S. L., Thiry, H., & Liston, C. (2012). The impact of a university-based school science outreach program on graduate student participants' career paths and professional socialization. *Journal of Higher Education Outreach and Engagement* 16(2), 47-78.

Snow, S. E. (2013). *Memories of GAMES: Measuring the Long-Term Impacts of After-School Museum Programming on Girls' Attitudes Towards Science*. Unpublished masters thesis, University of Colorado Boulder.

Thiry, H., Laursen, S. L., & Hunter, A.-B. (2008). Professional development needs and outcomes for education-engaged scientists: A research-based framework and its application. *Journal of Geoscience Education* 56(3), 235-246.

Appendix A: Post-program questionnaire for girls

1. Since coming to the museum for GAMES, I like science:

   More than I did before:_____

   Less than I did before:_____

   The same as I did before:_____

2. Please draw a line to connect the type of science with what it studies:

Archeology	Fossils
Botany	Animals
Entomology	Ancient Humans
Paleontology	Insects
Zoology	Plants

3. What did you like best about GAMES?

4. Pick one tool that you were given and tell me how you will use it in the future.

5. What did you like least about GAMES?

6. What would you like to be when you grow up?

Girls at the Museum Exploring Science (GAMES) Survey

Thank you for answering our questions about girls and science.
It should only take about 5 minutes. There are no right or wrong answers and if you do not like a question, you do not need to answer it.

When you were in 4th or 5th grade, did you participate in the Girls at the Museum Exploring Science (GAMES) program?

Yes
No
I don't know

If yes, please share something that you remember about GAMES.

_____

_____

Have you participated in other science programs outside of school?

Yes
No
I don't know

If yes, please list the programs you have participated in.

_____

I like science...

more than when I was in elementary school
less than when I was in elementary school
the same as when I was in elementary school

	Language Arts	Secondary World Languages	Math	Science	Music	Art	Social Studies	History	Physical Education	Health	Technology	Theater	Other
My favorite subject in school is...													
My *least* favorite subject in school is...													

For the following questions, please check all that apply.

	Language Arts	Secondary World Languages	Math	Science	Music	Art	Social Studies	History	Physical Education	Health	Technology	Theater	Other
Some subjects that I am good at are...													
Some subjects that are hard for me are...													

In the future, I plan to (check all that apply)...

graduate from high school
complete a GED
go to college
continue school after college
I am not sure

I would like to have a career as a(n)...

_____

In addition to these surveys, we will be talking with groups of GAMES girls to learn more about their experiences with the program. If you participated in GAMES, please consider participating in a group discussion. To thank you for helping in these discussions, we will provide you with lunch and a $5 gift certificate.

May we contact you about participating in a group discussion?

Yes
No

If yes, please provide a name and current school as well as your preferred contact information.

Name _____
Current School (Fall 2012)_____
E-mail _____
Telephone _____
Address _____

Thank you for your help! Please use the enclosed pre-paid envelope to return this survey along with the signed "Parental Permission & Student Assent" form by August 10th.

Appendix C: Interview Questions for Girls and Teachers

Focus groups for girls were open-ended and semi-structured. The use of guiding questions ensured coverage of relevant themes and issues, while allowing for spontaneous comments to emerge. Prompts included:

(1) Did you ever talk to people about GAMES? What did you tell them?

(2) What would you change about GAMES? What would make GAMES better?

(3) Did you ever use any of the tools in your toolkit?

(4) Do you go to any other out of school science programs?

(5) Do you like science more than, less than, or the same as you did in elementary school?

(6) What's your favorite subject in school?

(7) What do you want to be when you grow up?

Questions for teachers and chaperones were focused on their observations of students involved in the GAMES program. The semi-structured, open-ended interviews were guided by prompts such as:

(1) Describe your involvement with the GAMES program.

(2) What do you think GAMES participants get out of the experience? Why? Can you think of any stories or incidents with students that illustrate this?

(3) Did you notice any changes in the students who participated in GAMES either during the program or after its completion? How long did these changes last?

(4) Were there any noticeable differences between the program participants and non-participants within your class?

(5) What would you have liked your students to get from GAMES that you feel was lacking?

(6) What kinds of students do you think benefit the most from GAMES? Why? How do you select students to participate?

(7) How could the GAMES experience be improved for you or your students?