The Handbook of Broader Impacts
Volume editors: Van Egeren, L.A. & Renoe, S.
Transformations in Higher Education: The Scholarship of Engagement book series
Publisher: Michigan State University Press
Publication date: 2022

Chapter 25:

Creative Evaluation to Improve your Programming - Practical Tools and Tips for Evaluating Broader Impact Programs

Anne U. Gold[1]; Stacey Forsyth[2]; Alexandra Rose[2]; Kathryn Boyd[1]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder
2CU Science Discovery, University of Colorado Boulder

Abstract

Are you in the process of developing Broader Impacts for a grant proposal? Or perhaps you've been funded and are now planning or implementing the education and outreach activities you proposed? Collecting feedback from potential and current program participants is critical to planning and implementing impactful and effective outreach and engagement. In this chapter, we describe the goals of program evaluation and share examples of evaluation strategies for various Broader Impact events or activities. We describe the steps for successful formative and summative evaluation designed to produce actionable results, whether you want to improve your program or obtain the data you need to report program impacts to funders or other stakeholders. We provide examples of evaluation strategies that can be used to assess different types of Broader Impacts events or programs and link to an evaluation toolkit that describes many low-cost evaluation tools. Although some Broader Impact providers may elect to work with professional evaluators to design and implement their evaluation, here we focus on low-cost evaluation approaches that can provide useful insights about programmatic strengths and areas for improvement. We discuss the benefits and challenges of different evaluation techniques and provide tips and lessons learned about effective evaluation strategies. We also discuss ways to find and work with an external evaluator and discuss important considerations such as whether you should request approval from an institutional review board for human subjects research and how to write an evaluation plan for a proposal. We provide various resources to learn more about different evaluation methods.

What is program evaluation and why is it important?

Do you wonder how you'll know whether a program you're proposing will actually achieve the goals you have in mind? Or maybe you feel good about an activity you've implemented, but you'd like to get some feedback on how to improve it if you do it again? Measuring a program's impact and gathering feedback from participants on what worked well and how the program could be improved is critical for strengthening the quality of any Broader Impacts activity. In fact, reporting of impact metrics is often required by funding agencies. The National Science Foundation (NSF) defines evaluation as a "*systematic investigation of the worth, merit, value or success of a program or project*" (Frechtling, 2010). Overall, program evaluation is an important component that can support all stages of a successful project, from program design to final reporting (Patton, 2014; Friedman, 2008).

Evaluation can be used even before you start implementing a program. This sort of evaluation is called **front-end evaluation**, and it allows you to collect information--from prospective participants or representative individuals--that you can use to inform your approach to the project. Front-end evaluation might take the form of a short registration survey or a brainstorming session during a kickoff event. Listening to participants is critical to designing programs that feel engaging and relevant to the target audience. Front-end evaluation can often be informal as it is used less for reporting and more for your own purposes. Perhaps you are planning to develop a curriculum for a summer program for high school students. If you have a chance earlier in the year to speak with a group of high school students, you can use this opportunity to determine their level of familiarity with your topic by asking them some topic questions or having them participate in a trivia game that will reveal their knowledge and interest.

Formative evaluation, conducted while your project is ongoing, enables you to make iterative improvements throughout the life of the project. This might involve informal check-ins or 'exit tickets' with program participants or include evaluation activities that are embedded into the program. When done effectively, embedded evaluation can serve multiple purposes: generating useful information and feedback while also serving as a meaningful component of the program. For example, a reflection activity can provide closure to any type of Broader Impact event while also providing an opportunity for participants to share what worked well (or not) and which aspects of the program were most memorable or impactful. The results of formative evaluation are intended to be used and implemented in a timely fashion, so be sure to consider how long it will take you to analyze the results of the assessment tool you deliver.

Summative evaluation provides insights about a project's outcomes. Summative evaluation might engage participants in reflection on the overall program, focusing on what your audience has learned or how their attitudes about your topic have changed. Summative data are usually collected once the program has ended and these data allow you to document results and program impact. Program reports and publications usually include summative evaluation data as evidence of a program's success (Frechtling, 2010). When planning summative assessment, it is important to consider all of your project's goals and audiences, potentially designing different tools for assessing them.

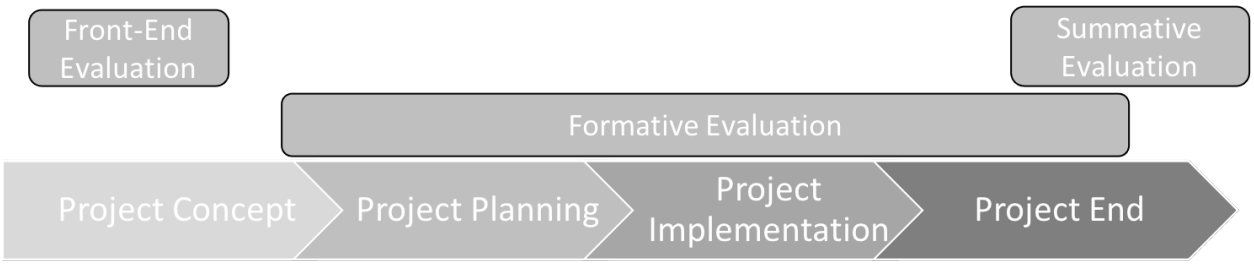

Figure 1. Illustration of evaluation stages throughout the different project implementation phases.

How can evaluation help your program?

- ✓ Understand, verify or increase the impact of a program
- ✓ Ensure that a program is meeting its goals
- ✓ Identify program strengths and weaknesses to improve a program
- ✓ Verify that you're doing what you think you're doing
- ✓ Produce data that can be used for publications or future proposals
- ✓ Enable you to share outcomes with funders or community stakeholders

How do I get started?

- ❑ Identify target audience(s)
- ❑ Identify goals of your Broader Impact activity (outputs, outcomes)
- ❑ Identify appropriate evaluation strategies and techniques
- ❑ Secure Institutional Review Board (IRB) approval (if necessary, see below)

When you're planning your Broader Impacts activities, it's important to first consider these two questions: *Whom do you want to reach with your work?* and *What do you want that audience to take away from the interaction or program?* Determining your **audience** may help you narrow down the format of your outreach work. For example, if you're interested in reaching K-12 science teachers, designing a museum exhibit on a University campus may be less useful than developing online educational resources or providing opportunities for students to connect with scientists.

When scientists have a particular audience in mind, such as schoolchildren or retirees, but don't have a clear picture of what kind of programming they want to offer, it helps to focus on their **outreach goals** to decide what programming would be best. Goals are the metric by which your program's success will be evaluated, so it makes sense to spend time clearly articulating and refining your goals. Goals are usually a broad statement about the overall expectation of what will happen as a result of your program. Goals are often phrased with action verbs ("increase..", "inspire…", "raise awareness..") and should be measurable (Fig. 2). Keep in mind that the format and duration of your program will affect the outcomes you may expect to see (Cappelli et al., 2019). For example, programs with short durations should focus on short-term outcomes, such as awareness, interest and attitudes, whereas longer-term programs (more than 60 hours) may want to focus on longer-term outcomes, such as academic outcomes and interest in related careers (Wilkerson and Haden, 2014).

Evaluators often use the acronym S.M.A.R.T to help develop project goals. To develop S.M.A.R.T. goals you need to keep in mind the following characteristics and guiding questions:

- ***S****pecific:* What exactly are we going to do for whom?
- ***M****easurable:* Is it quantifiable or describable and can we measure it?
- ***A****ttainable:* Can we do it in the proposed timeframe with the resources we have?
- ***R****elevant:* Will it have an effect on the desired goal/strategy?
- ***T****ime-based:* When will this goal be accomplished?

For more information on SMART goals and how to develop them see this CDC evaluation report (2017) or Chapter 4 in Poister (2008).

Figure 2. Illustration of SMART project goals (modified from Dungdm93, 2016).

Goals can be broken down into objectives, outputs and outcomes in order to make assessing them more tractable. Goals tend to be visionary and aspirational, whereas outcomes are more targeted and specific, describing the changes you expect participants to experience through your program (CAISE 2011). Your **objectives** are the results you wish to achieve and the manner in which they will be achieved. Stating your objectives for an activity might help you focus on the specific content or skills you wish to teach. You will usually define more than one objective to meet a particular goal. Just like goals, objectives should be as specific and detailed as possible. For example, if your goal is to improve students' data science literacy, your objectives may be to have students learn how to interpret graphs, design and implement a plan for data collection and analysis, and present their findings using a chart or infographic to highlight compelling results. **Outputs** are the specific products of your work, such as the number of participants in your program or the number and type of educational resources developed and shared. **Outcomes** are the immediate or long-term changes you will monitor in order to assess your progress towards achieving your objectives. Outcomes may be changes in knowledge, skills, and behavior. Outcomes have clear and measurable criteria, e.g., 83% of families reported accessing information about real-time air quality in their community after the program as compared with 45% of families who had accessed the information prior to the program.

A useful approach for outlining Goals, Objectives, Outputs and Outcomes in an organized and intentional way is to use a Logic Model (Fig. 3). Logic models are a visual representation of the work you have done to design your project. The logic model shows through text and graphical connectors how the project activities result in your intended outputs and outcomes. Logic models are flow diagrams that list a *problem statement* (e.g., lack of diversity in science workforce), the *inputs* you are providing for the project (e.g., team experience, funding), the *activities* that are planned (e.g., organizing science cafes), the *outputs* (e.g., science cafe events) and the intended outcomes, usually separated into *short term outcomes* (e.g., participation in a discussion about a science topic) and *long-term outcomes* (e.g., increase in student interest in science). Developing a logic model will be a helpful exercise for the design of your project. Including even a basic logic model in a proposal can be a powerful tool for demonstrating clear intention for the Broader Impact activities and an excellent method for organizing your own thoughts and goals.

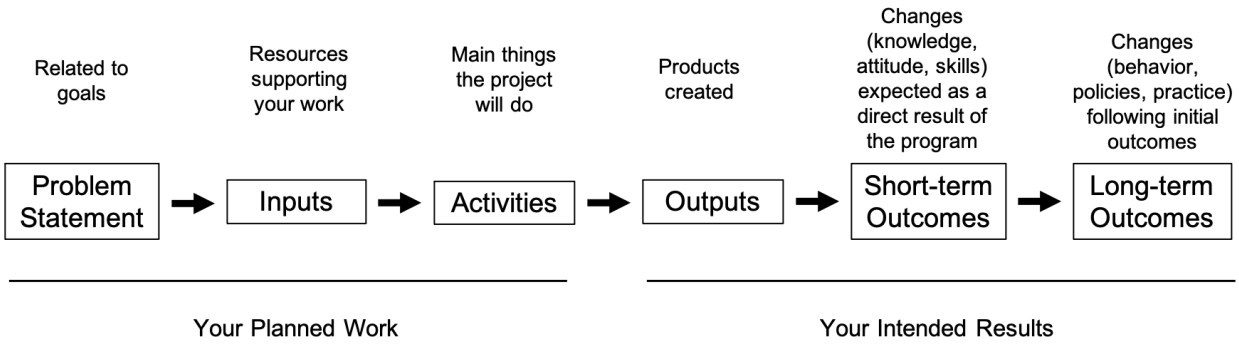

Figure 3. Basic Logic Model. Adapted from the W.K. Kellogg Foundation (2004) & EvaluATE (2021)

Once you have identified measurable goals and outcomes for your Broader Impacts project, it is time to decide which **evaluation strategies and tools** are appropriate for your audience and your activities. The answer will depend on a number of factors, including the amount of time you spend with your participants (e.g., a short webinar presentation will be evaluated differently than a 1-week summer camp), the experience your team has with evaluation, the program setting (e.g., virtual or in-person environment) and your access to materials and tools. Asking yourself about the nature of the outreach, what artifacts (e.g., drawings, maker space products, videos), if any, will be created by participants and what kind of interactions seem appropriate will help identify which evaluation tools will be most effective. If you are using survey tools, you need to research which context they were developed for, what audiences they were tested with and what are appropriate data analysis strategies to select the appropriate survey instrument for your project.

How much evaluation is needed?

Evaluation is possible on a shoe-string budget if conducted by the project team, but large programs often include comprehensive data collection by external evaluators. External evaluators can provide an objective view and avoid bias in the analyses and reporting. Comprehensive evaluation designs often work backwards from the project goals to define outcomes and impact and select measurement tools to collect evidence to determine if program goals were reached (see logic model discussion above and Fig. 3). If the program budget doesn't allow for comprehensive, external evaluation, the project team can integrate evaluation activities into the program design and allow for data collection on a small budget. It is important to scale the evaluation plan to the size of the project and to the goals of the data collection. You might want to ask yourself if you are primarily collecting data to improve your program, or if you are, for example, interested in publishing your findings. Depending on the answers to these questions, you will need to develop different evaluation plans. Publications require rigorous data collection and a thoughtful research design.

> More information on how to find and work with an external evaluator can be found here: https://www.informalscience.org/evaluation/working-evaluator

Creative Approaches to Evaluation of Broader Impact Activities

One challenge in developing an effective evaluation plan for outreach efforts is to select the most appropriate evaluation tools and approaches for your program. We find that it is critical to decide what is absolutely necessary to know. It is tempting to ask too many questions but especially in informal learning environments it is important to keep the evaluation activity fun and short. Nobody wants to feel like they

have to complete another quiz or school-like assessment. Finding creative ways to gather feedback is important for the quality of the data as well as for participants' morale. Identifying playful ways of collecting data from participants can yield valuable feedback for program facilitators while ensuring that the process is interesting and enjoyable for participants.

If you drafted your goals and outcomes as described above, you are ready to develop your evaluation plan. An evaluation plan is a road map through your program activities, the path for iterative improvements and the guide to measurements of outcomes. Many playful evaluation tools can be embedded in the activities and serve as icebreakers, consensus-builder, or reflection activities. For example, you might have participants split out in groups by experience or interest in a maker space workshop and have them put their workshop goals on sticky notes on large poster boards (front-end evaluation). Right before a lunch break you might have participants write on an exit ticket what their three favorite activities of the morning were and what could be improved (formative evaluation). At the end of a workshop, you might ask participants to record a short video reflection using a tool such as Flipgrid to answer some broad questions that connect to the program outcomes (summative evaluation).

Table 1. Examples of Evaluation of Broader Impact Activities. For a more detailed overview visit the evaluation toolkit: https://www.colorado.edu/researchinnovation/broader-impacts-project-evaluation

Tool	Description	Setting/ Learning Environment	Required Materials or Technology
Focus Group	Focus groups typically include ~3-12 participants who provide their ideas and insights about a particular topic. Focus groups can be used as a part of front-end evaluation, to learn about a community's needs or interests before beginning a project, or as summative evaluation, to collect qualitative data that reveals insights relating to a program's impact.	Any	Minimal: paper for notes or device for recording the session
Sticky Note Feedback	Asking participants to write a quick response (e.g., 'What will stick with you?') or draw an emoji can be a quick and easy way to assess which parts of an experience most resonated with participants.	In-person programs; Virtual programs	Minimal: sticky notes & poster paper/board or virtual board (e.g., Jamboard)
Exit Tickets	Exit Tickets usually include a short reflection question at the end of a program. They can help assess overall impressions of or learning through a program, project, or lesson. Exit tickets are designed to be collected before participants depart.	In-person programs; Virtual programs	Minimal: paper or index cards
Graffiti Wall	Graffiti walls enable participants to share brief comments or questions and create a visual representation of feedback and overall engagement. Online whiteboards (such as Jamboard) can support graffiti walls in remote (virtual) programs.	In-person programs; Virtual programs	Minimal: sticky notes & poster paper/board or virtual board (e.g., Jamboard)
Concept Map	Concept Maps prompt participants to present their knowledge or understanding in a visual	Any	Large paper, colored pencils or markers, or

	way by writing or drawing what they think about a topic and drawing connections or links between related ideas. This approach can be used to gauge participants' familiarity with a topic at the beginning of a program, as well as to assess how their understanding changes over time.		on-screen equivalent
Snapshot Interviews	Snapshot interviews allow participants to provide feedback in a short and efficient way. Typically consisting of just 3-4 questions, they take only minutes to complete, making them easier to implement as a quick feedback tool.	Any	Minimal: paper for notes or device for recording the interview
Video Reflections	Video reflections recorded by participants can help provide a deeper and more nuanced understanding of participants' experiences. Participants record a short response (~1-5 minutes) to customized questions or prompts. Online tools such as Flipgrid enable facilitators to readily collect and organize short video reflections from participants.	Any	Device to record audio and/or video (e.g., phone, tablet, computer with webcam); program to organize videos (e.g., FlipGrid)
Maker Products/ Artifacts	Makerspaces are unique learning environments that require a different approach to evaluation. Making activities are typically open-ended with opportunities for student voice, creativity, problem solving and iteration. Evaluating participants' artifacts, potentially in combination with the Maker(s) describing their design decisions, may provide a deeper understanding of participants' learning than other evaluation approaches, such as surveys.	Makerspaces; STEM programs	Variable, depending on project
Online Polls	Online polling tools make it easy to conduct polls in real time with participants, either in person or online. Results can be shared immediately, as charts or word clouds, or downloaded for further analysis.	Virtual programs	Computer, tablet or cell phone

What methods are used in Broader Impact Activity Evaluation?

Surveys are one of the most commonly used tools in evaluation. They are useful for efficiently collecting and quantifying information from Broader Impact program participants. We have used surveys to address various goals, including exploring topics that participants are interested in, gathering feedback about an event, measuring audience characteristics, understanding audience needs or interests, and quantifying changes in participants' knowledge or behavior before and after programming. We will highlight advantages and disadvantages of surveys as well as some alternatives, including simple, inexpensive evaluation techniques, after we take you on a short tour of survey-based evaluation. Survey questions are often at the center of alternative evaluation activities (e.g., prompts for video reflection or focus groups, questions for voting activities), so understanding the basics of surveys is critical.

What are the advantages and disadvantages of using surveys?

Advantages of surveys are numerous - they are relatively inexpensive to facilitate, are good for gathering descriptive data, can be used to gather information on a variety of topics with relatively large numbers of participants, and results can often be analyzed easily with software (Frechtling, 2002). However, there are a number of disadvantages too. Surveys may remind people of school and feel like a test and thus not fit in the context of the program you are offering. Surveys are also not accessible for non-readers/non-writers like young children. We have also found that surveys don't always provide much context to the responses provided (e.g., why did one participant connect with a certain aspect of the program while another didn't?) and often don't usually encourage or facilitate deep reflection about the program impact. Another consideration for all assessments, but specifically for surveys, is that self-reported information can introduce bias from participants (e.g., reporting their level of awareness about a topic they may have limited knowledge about). Survey responses can also be skewed by the desires for social approval (e.g., wanting to please the facilitator) or stereotype threat (APA 2006; e.g., if participants are asked about their gender and race/ethnicity before being asked what they think of an engineering exhibit, research has found that participant responses can be influenced by stereotyped views of who does engineering). In general, surveys are best used for capturing the 'big picture' on a project.

Tip: Different survey platforms offer different ways to collect anonymous data. In general, Qualtrics and Surveymonkey (to name some of the most common tools) offer easy ways to collect data anonymously. However, many students and teachers are very familiar with Google forms and may prefer surveys in Google as they can possibly be better integrated in learning platforms that are used as part of the activity. Your home institution may offer a free institutional license for Qualtrics.

What kinds of questions should be asked in a survey?

Survey questions should be designed to address the evaluation goals through either qualitative or quantitative data. **Closed-ended questions** with selectable options are typically used to collect quantitative data, which can be more easily analyzed with statistical software packages. However, the data you are collecting will only be as good as the answer options you are providing. Also giving participants a limited number of answer choices can lead to response bias, if their response is not reflected in the choices. **Open-ended, text entry questions** allow participants to provide more context and other relevant information but can take much longer to analyze using qualitative thematic analysis methods. **Likert scale questions** typically ask for responses on a scale, such as the level of quality of a program, how much participants agree with a statement, or how often they engage in a certain behavior. These questions collect personal reflections and are a common feature of surveys. Online survey tool, such as Qualtrics and SurveyMonkey, usually provide support in finding appropriate answer options for your Likert question. You may also consider giving both a pre-program survey and a matched post-program survey. Such a **pre-post survey design** allows you to identify shifts in responses from before to after the program. An alternative to the pre-post survey design is the design of **retrospective questions** on the postsurvey. For example, you might ask respondents to rank their opinion about statements indicating how they felt before the program and how they feel about the same statements after the program. This design will provide information on the relative gain a participant perceived over the course of the program (Vaske, 2009).

How do you create a survey?

Survey development can be daunting and if you want to embark on the adventure, we encourage you to work with experienced colleagues to help create and review questions. The process requires research on relevant literature or other background knowledge and identification of the goals for the data collection (Vaske, 2009). Survey developers ensure the questions make sense and are appropriate for their

intended audience, through piloting testing of the surveys. Also remember to test how long it takes to complete a survey so you can let participants know how much time they will be spending.

Surveys can be conducted online, by phone or mail, or in person with pen/pencil and paper. Participants may be hesitant to invest the time needed for a survey, which can lead to low response rates. If your program has the funds, incentives such as gift cards or prize drawings provide an effective way to increase the survey response rate. A $5 gift card goes a long way for a short survey, but you may want to offer a larger incentive for an hour-long, complicated survey. Response rates will also decrease when participants encounter issues accessing surveys or if they don't understand the questions. Using QR codes and shortening the web address for online surveys with programs like TinyURL or Bitly can help respondents easily participate.

Given the significant time and effort required to develop surveys, you could consider using existing survey instruments that have been tested for reliability and validity in a similar setting. There are many repositories of surveys that can be used to measure attributes such as science interest or science identity that are commonly used to measure Broader Impact project outcomes (see insert box below).

Before you collect data from participants, we strongly recommend that you check with your Institutional Review Board (IRB, see more information below), if you are required to collect permission for the evaluation activities.

What kind of data will I collect?

Evaluators differentiate between **quantitative data**, such as rankings on survey responses or votes contributed through sticky notes, clickers or colorful cards, and **qualitative data**, such as open-ended survey responses or video reflections. Quantitative data may be easier to report on but may not provide as much insight into participants' experiences (Patton, 2008). Qualitative data, on the other hand, is more subjective but can often provide rich insight into participants' perspectives. Qualitative data can also be used as testimonials to illustrate what participants thought about the project. To analyze qualitative data, the evaluator needs to group responses in different thematic groups ("codes"). The frequency with which a certain code was used by participants can be reported, quantified and analyzed. More information on how to conduct and the theoretical basis for thematic analysis can be found in psychology literature, including Braun & Clarke (2008). They describe several steps to generating and finalizing themes or codes: 1) familiarizing yourself with data, 2) generating initial codes, 3) searching for themes, 4) reviewing themes, 5) defining and naming themes, and 6) producing the report.

Repositories of Evaluation Tools and Instruments:

- Activation Lab - The ActLab website includes many evaluation instruments to measure constructs such as science learning activation, engagement, and scientific sensemaking.
- Assessment Tools in Informal Science (ATIS) - ATIS hosts a searchable database of evaluation instruments used in informal STEM learning settings.
- STELAR (instruments used in NSF ITEST projects) - The STELAR center curates a collection of evaluation instruments used in STEM education and career development contexts.
- CAISE (evaluation tools & instruments for informal science) - The CAISE center curates a collection of evaluation instruments used in informal STEM education.
- Developing, Validating, and Implementing Situated Evaluation Instruments (DEVISE) - The Cornell Lab of Ornithology developed a series of instruments to measure individual learning outcomes from participation in citizen science such as interest, motivation, self-efficacy, and skills.
- Online Evaluation Resource Library - This NSF-funded library includes tools and resources to design, conduct, document, or review project evaluations.

Going Beyond the Survey

As described above, surveys do have some limitations and may not always yield a full understanding of participants' experiences. Other methods for gathering participant information and feedback include interviews, focus groups, observations, content assessments, participant artifacts, and participant reflections collected through various means (e.g., video, voting methods, sticky notes on whiteboards, postcards, reflection activities, etc.). See a compilation of various evaluation methods here: https://www.colorado.edu/researchinnovation/broader-impacts-project-evaluation.

Interviews and focus groups enable an interviewer to explore topics in greater depth, gauge participant feedback during face-to-face interactions, and allow flexibility or follow-up questions to clarify responses. However, interviews and focus groups are time-intensive, including both time spent in the interview as well as the time required to transcribe and analyze the large amount of qualitative data generated. Interviews and focus groups also require well-trained interviewers because social interactions with interviewees can introduce bias, and the flexibility of the questioning requires someone to maintain consistency, or account for any differences through the analysis (Frechtling, 2002).

Depending on the nature of your Broader Impact program, **products or artifacts** created by participants during the program's activities may also serve an evaluation purpose in addition to the program purpose. They can provide you with insights about the program's effectiveness and impacts; for example, if youth participate in a film-making workshop about a topic relevant to your Broader Impact effort, they might create a mind map as a brainstorming document on a large sticky note and later create a visual representation of the storyline. These documents provide some insights into the process and understanding and can be used to identify different steps in the collaborative process.

Informal education settings such as museums often use quick and easy data collection methods because participants don't want to spend time taking surveys. **Sticky notes** can be a quick and engaging way to brainstorm participant ideas or feedback. Each participant is asked to respond to a prompt on a sticky note and post it on a large poster board or white board. The boards can be used to organize participants' ideas through timelines, Venn diagrams, and charts or graphs, to name a few. CU Boulder's Teen Science Café, for example, uses sticky notes to gather teens' questions about a topic before an event and then provides a "What Stuck with You?" board to collect their immediate impressions of a café event as participants leave. Writing a short response on a sticky note is quick and easy but provides some direction to guest scientists about what teens are interested in and provides program organizers with useful information about what was most interesting or memorable for participants. We also have used a **"bucket" survey method** in the local planetarium, where, under the supervision of staff, a multiple-choice question (e.g., "Which planet do NASA satellites study the most?") is presented on a white board and audience members are invited to deposit a token (poker chip) in the box that illustrates their response (e.g., "Earth", "Mars", "Jupiter", "Saturn") to the question as they exit the planetarium.

Another way to collect feedback in a creative way is to ask participants to **draw a picture** based on a prompt. The most famous example is the "Draw a Scientist" prompt to collect information on the associations that participants have with scientists (Chambers, 1983). The information might be insightful if, after the completion of a program, certain characteristics of the drawing change. For example, if drawings of scientists before the program show gray-haired men in lab coats and after the program the participants draw young women. A related method is the **personal meaning mapping** approach (Falk, Moussouri, and Coulson, 1998), where a drawing prompt that is relevant to the program is given to participants (e.g., "What does water mean to your community?") and the participants create a concept map or brainstorming drawing about the topic.

There are also many other methods used in evaluation. We use **content assessments** to consider gains in learning or understanding of certain topics. Such content assessments can come from classroom assessments professors may already have. **Observations or video recordings** of program activities or group interactions can be analyzed for participant behaviors or actions. **Social network analysis** can be used to consider participant interactions or map network connections (Carrington et al., 2005).

Resources to learn more about developing surveys, conducting interviews, and other data collection or analysis methods:

- Training through your local Education Departments or Centers for Teaching & Learning
- Online courses through the American Evaluation Association (AEA)
- Hughes et al (2016) - article exploring survey demographic questions
- Vaske (2009) - book about survey research and survey analysis

Creative Evaluation in Action

We are providing a few examples to illustrate how Broader Impact activities can be evaluated.

Creative Reflection

An Aerospace Engineering faculty received an NSF CAREER award for her research involving remote sensing of atmospheric change. Her passion for involving non-traditional and English Language Learners in STEM fields led her to partner with a local K-12 outreach group on her campus to offer a series of workshops for high school students around remote sensing technology. She and her graduate students led hands-on workshops where students built and programmed their own remote sensing cameras and used them to collect data on the environment around them to learn about remote sensing technology. In addition to using a traditional survey to assess student learning, she also developed a creative reflection tool. The tool was designed to explore how students connected to the workshop with different senses: their Heads (Academic Skills and Knowledge), Heart (Social Emotional-Learning and Self-Efficacy), Hands (Action and Service) and Feet (Connection to Place). The reflection prompted students to write about, draw, or photograph something illustrating each of these connections with the workshop - their products revealed meaningful discoveries about the student experience.

Designing evaluation for meaning making

For several years, an interdisciplinary team of faculty researching an NSF-funded Critical Zone Observatory (CZO) in Colorado collaborated with STEM educators to share current CZO science with elementary and middle school age audiences. A major component of this Broader Impacts effort involved a one-day Earth system science workshop delivered numerous times at schools throughout the state. The short duration of the program posed a challenge for evaluating the program's effectiveness, as program facilitators didn't want to give up time for evaluation activities like surveys. In order to balance these competing needs, program coordinators developed a culminating activity in which students worked in teams to design and present a poster that highlighted key things they had learned during the workshop. Because the task was integrated into the workshop itself, students viewed it as a fun, collaborative activity, rather than as an additional evaluation task. They enjoyed the opportunity to creatively express what they had learned and share their ideas with their peers and teachers. At the same time, however, this exercise, combined with a short survey issued to teachers, enabled the team to identify lingering misconceptions and gauge the workshop's effectiveness in meeting the project's learning goals.

Figure. 4. Student-designed poster highlighting key concepts learned during an Earth system science workshop.

Envisioning a Teen-Friendly City

As part of a larger community initiative to design a child-friendly city, CU Boulder researchers partnered with the university's Teen Science Cafe program to gather teens' input about the kinds of places and resources they wanted to see in their city. The team met with several small groups of teens and to spark conversation, they provided each participant with a mix of randomly selected materials, including recycled objects, small toys and a mix of craft supplies. Teens were tasked with designing a physical representation of a space they'd like to see in their community, using only the miscellaneous materials they had snagged from a bag. The creative design task provided time for teens to work alongside each other in a social setting as they built a physical representation of an imagined space. After a short 'design and build' phase, teens shared their designs with researchers and their peers, using their physical models to help communicate their designs. Although this exercise was completed in a short amount of time (~20 minutes per group), it prompted rich conversation and provided insight into the kinds of spaces teens wanted to see in their communities. Additionally, the playful nature of the design task - in which they were building parks, playgrounds and hang-out spaces with materials like pipe cleaners, toys and beads, helped make the experience fun for participants, rather than a chore, leading to a productive conversation with an audience that is frequently reluctant to communicate

Evaluation in Makerspaces

Makerspaces are creative, open-ended environments in which participants have a lot of flexibility in what they design and how they do it, and the focus is on the process rather than the product. By their nature, these spaces differ in important ways from other learning contexts, which can pose some unique challenges for evaluating programs in these settings.

One Broader Impacts project, connected with an engineering faculty member's research, engaged youth in Makerspaces to design books, games and other products to be more accessible for audiences that are blind or visually impaired. Because surveys failed to capture some of the program's most salient outcomes, project directors shifted to other approaches. A simple rubric based on key design principles enabled the team to evaluate participant-designed artifacts. In addition, participants created short video reflections in which they shared their designs, explained their rationale for the different features they included, and discussed how they overcame certain challenges throughout the design and fabrication experience. Coding and analysis of these video reflections provided the evaluation team with an enhanced

understanding of the program's impacts on participants as compared to surveys, which failed to fully capture these impacts.

Evaluating Broader Impacts in Virtual Environments

Delivering Broader Impacts programming in virtual environments poses a new challenge in terms of how to evaluate program impacts. Unlike in-person Broader Impacts programs, remote programs delivered via online platforms such as Zoom can make it harder to gauge participant engagement. Fortunately, a variety of online resources now make it easier to collect participant feedback and work products online.

A one-week summer program for high school students focused on a Computer Science faculty's research, aiming to increase participants' awareness of artificial intelligence (AI) and the ethical issues related to AI-powered technologies. In addition to a short pre-post survey, the team used a variety of other approaches, including embedded activities, such as a quick drawing exercise and multimedia design projects, such as comics, videos and chatbots, to evaluate teens' understanding of AI and related ethical issues. The team used Padlet, which functions as an online bulletin board, to collect students' digital drawings and other products. Review and evaluation of student artifacts provided new insights about the ethical issues that teens were most concerned about, as well as changes in participants' understanding of key issues as the program progressed.

Culturally responsive evaluation

Depending on your participants or your regional setting you may want to consider the cultural context for your evaluation. From our experience, many Broader Impact programs are focused on engaging or supporting underrepresented groups in science. Therefore, the success of a Broader Impact evaluation is dependent on the evaluator's ability to identify and respond to the cultural setting of the program and consider the participant's cultural context and community (Frierson, Hood, Hughes, & Thomas, 2010). Culturally Responsive Evaluation (CRE) has been described as an *approach* to evaluation, rather than a specific methodology (Casillas, Hopson, & Gomez, 2015). That means CRE does not have a certain set of activities that are taken to perform an evaluation. Rather, there are several principles that need to be considered when taking a culturally responsive approach to evaluation. Casillas et al. (2015) laid out these steps to take towards a culturally responsive evaluation as follows:

- Understand and recognize the larger context of the program and evaluation
- Design the evaluation with participants in mind
- Allow for self-determination by the participants
- Build trust and facilitate communication with participants
- Understand the evaluation audience and make the evaluation accessible
- Examine evaluator attributes which may affect efforts.

CRE can be very time and energy intensive, but when working with diverse audiences we have found that it was important to spend the time to understand the cultural context to conduct a meaningful evaluation of program activities. CRE is always an iterative process and requires that the evaluator is willing to adapt the methodology and embed it in meaningful cultural practices or contexts. Different organizations offer professional development workshops that can help you learn more about CRE approaches (see inset box). An example from our work of using a CRE approach in program evaluation is an NSF AISL-funded project in which we developed a traveling exhibit and programs for rural libraries around water topics. We developed a needs assessment survey to gather information from the library audiences about the topics and type of programming they would be interested in as a front-end evaluation activity. The survey data showed that the respondent demographics were not reflective of the region. Working with project partners and advisors, we determined that interviews were more likely to be an effective way of collecting information from the audiences that were under-represented in the surveys and it would also allow participants to provide context about their responses. We deliberately recruited

community members from diverse groups and took the time for individual interviews. During the interviews, the participants were willing to share their experiences and expectations through storytelling and perspective sharing that our structured surveys did not allow for. Although the process took longer, we learned much more and were able to inform the project design with the voices from our target audiences in our needs assessment.

How to write the evaluation section in your proposal?

A competitive evaluation plan helps pull all parts of a proposal together and will strengthen and connect it for the reader. It also shows that you care about improving the program and learning. Your evaluation plan will look very differently for projects that have a large dedicated evaluation budget in comparison to proposals that integrate evaluation activities as a small component of the work. Of course, the evaluation plan needs to match the solicitation and the scope of proposed activities. Here are some tips from our team about writing evaluation sections for your Broader Impact project.

- Identify who will be doing the evaluation in your evaluation section. An external evaluator? Your team? Name the person and highlight their qualifications in the evaluation section.
- Match the evaluation plan to the project goals, objectives, outcomes, and activities. Don't use standard text that is not aligned with the proposed work. Consider using a Logic Model to organize these thoughts.
- Design your evaluation plan to collect information about what is working and where adjustments and improvements are needed. Show that the evaluation will inform the project design and revisions.
- Draw from established evaluation practices (see infobox for databases) to create your evaluation plan.
- Be realistic about the evaluation you are proposing - the budget you set aside should match the evaluation plan. Small evaluation budgets can only afford a small number of activities.

Teles (2020) summarized helpful hints and fatal flaws for developing an evaluation section for proposals, if you want to learn more.

What do I need to know about institutional review boards?

If you want to publish or report on your data, you may need approval from your local institutional review board (IRB) and to ask your participants to provide consent for participating in your study. The board works to protect the rights of participants, especially minors and other vulnerable populations. In general, any data collection on participants that will be used to generalize findings (e.g., publications) need to be approved by the board. On the other hand, formative evaluation data that is being collected only for iterative project revisions can be collected without requesting consent from the participants.

For IRB approval the board reviews a detailed research plan together with a description of the questions that you plan to ask and all recruitment materials. Usually, participants have to provide informed consent in order for an evaluator to use the data for a study. For minors to participate in an evaluation study, parents need to provide consent along with the youth's assent. However, collecting formative feedback to improve the program does usually not require IRB approval. Rules vary from institution to institution and so does the amount of time it takes for the board to review applications.

Wrap Up

We hope that this overview has sparked some good ideas and inspiration to try creative ways to evaluate your program. Evaluation has a place at all stages of the project. If your project design is guided by evaluation data about the interests, needs and expectations of your participants you will likely create a very impactful program. Iterative improvements throughout the project can only be made if you collect data to inform revisions and you want to measure impact at the end of the project to inform funders or future project design.

Don’t be intimidated to try out some creative ways of collecting feedback. You don’t have to develop long surveys; short and quick reflections will often provide helpful insights without interrupting the flow of your program. Keeping the evaluation short and sweet will also be a reminder to only collect the data that you need. Participants enjoy creative interactions, and each creative evaluation activity will likely help your participant reflect and get even more out of the program. Many of the examples we provided can be embedded in your programming.

If your goal is more than iterative revisions of your program, for example, maybe you want to publish findings about the program, you will need to develop a more extensive evaluation plan, likely apply for institutional review board approval and bring in a professional evaluator.

Evaluation is an invaluable tool to make your project successful. From the very beginning when you sit down and think about the alignment of project goals, activities and the impact of your work, being thoughtful about ways to evaluate your program will improve the project design and outcomes. Whether you are embarking on a large evaluation or a small-scale assessment, we wish you luck and encourage you to make use of the many evaluation resources out there.

Acknowledgements
The evaluation toolkit webpage was built with support from Casey Marsh and Gigi Yang. This chapter was supported in part by the NOAA Cooperative Agreement with CIRES, NA17OAR4320101 (Gold, Boyd).

References

American Psychological Association (APA) (2006). Stereotype threat widens achievement gap. American Psychological Association, https://www.apa.org/research/action/stereotype (accessed: March 2021).

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Cappelli, C.J., K. L. Boice and M. Alemdar. (2019). Evaluating university-based summer STEM programs: Challenges, successes and lessons learned. Journal of STEM Outreach 2: 1-12, July 2019.

Carrington, P. J., Scott, J., & Wasserman, S. (Eds.) (2005). *Models and methods in social network analysis* (Vol. 28). Cambridge University Press.

Casillas, W. D., Hopson, R. K., & Gomez, R. L. (2015). Making culturally Responsive Decisions in Evaluation Practice. In Christie, C.A. & Vo, A.T. (Eds.) Issues in Evaluation Use and Decision Making in Society (pp. 111-128). Information Age Publishing.

Center for Advancement of Informal Science Education. (2011). Principal Investigator's Guide: Managing Evaluation in Informal STEM Education Projects. Washington, DC: Author. http://informalscience.org/evaluation/evaluation-resources/pi-guide (accessed: March 2021).

CDC (Center for Disease Control) (2017): Evaluation Guide. Writing SMART objectives. https://www.cdc.gov/dhdsp/docs/smart_objectives.pdf (accessed: March 2021).

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

Dungdm93 (2016): SMART goals. Wikimedia. https://commons.wikimedia.org/wiki/File:SMART-goals.png (accessed: March 2021)

Evalu-ATE (2021): Logic Models for ATE Projects & Centers. https://www.evalu-ate.org/library/proposals/logic-models/ (accessed: March 2021)

Falk, J. H., Moussouri, T., & Coulson, D. (1998). The effect of visitors 'agendas on museum learning. Curator: *The Museum Journal*, 41(2), 107-120.

Frechtling, J. (2002). The 2002 User-Friendly Handbook for Project Evaluation. National Science Foundation.

Friedman, A. (Ed.) (2008). Framework for Evaluating Impacts of Informal Science Education Projects https://www.informalscience.org/sites/default/files/Eval_Framework.pdf (accessed: March 2021).

Frierson, H., Hood, S., Hughes, G. B., & Thomas, V. (2010). The 2010 user-friendly handbook of project evaluation. National Science Foundation, Directorate for Education and Human Resources, Division of Research and Learning in Formal and Informal Settings.

Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of Psychological Research*, *21*(3), 138-151.

Kellogg Foundation. (2004). WK Kellogg Foundation logic model development guide. WK Kellogg Foundation. https://www.aacu.org/sites/default/files/LogicModel.pdf (accessed: March 2021)

Patton, M. Q. (2014). Qualitative research and evaluation methods. Thousand Oaks. Cal.: Sage Publications, Edition 4.

Poister, T. H. (2008). Measuring Performance in Public and Nonprofit Organizations. John Wiley & Sons. p. 58-75. ISBN 978-0470365175.

Teles, E. (2020): 10 Helpful Hints and 10 Fatal Flaws: Writing Better Evaluation Sections in your Proposals. Blog post: https://www.evalu-ate.org/resources/10hint-proposals/ (accessed: March 2021)

Vaske, J. J. (2019). Survey research and analysis. Sagamore-Venture.

Wilkerson, S. B., & Haden, C. M. (2014). Effective practices for evaluating STEM out-of-school time programs. *Afterschool Matters*, 19, 10-19.