Evaluation Guide for Public Libraries

BY KATE HALEY GOLDMAN

HG&CO

VISITOR-CENTERED PLANNING, STRATEGY, AND EVALUATION

Table of Contents

©2021 Urban Libraries Council

Introduction: Why Evaluate?

The civil sector and nonprofit world have increasingly incorporated the practice of evaluation, often for a variety of internally- and externally-driven reasons: to create institutional change, to demonstrate the importance of specific programs or initiatives to funders and sometimes to simply demonstrate their own impact to the outside world. No matter the impetus, it is hard for program directors to refrain from feeling judged or resentful within the process of evaluation.

This training manual has been created so that you as a librarian feel empowered, rather than overwhelmed, when embarking on an evaluative process. Evaluation can, and should, be part of regular reflective practice — one that incorporates deep listening to your community and your community's needs. Evaluation doesn't need to feel overwhelming or frightening. It can, and should, be part of regular reflective professional practice. Evaluation can be done by those within the library and it doesn't have to be expensive or involve outside consultants.

Well-done evaluation serves internal library needs, helping the library achieve their goals, allocate scarce resources to where they are most effective, better understand their patrons and serve community needs. Evaluation can save libraries time and money by creating an environment where decision-making is based on evidence. Evaluation supports libraries, making the case for how libraries can effectively achieve their goals and connect with their patrons. It demonstrates the relevancy of libraries, showing their critical role within communities. Evaluation can improve how work is done within the library.

This guide is designed for library staff to gain familiarity with the topic of evaluation, and the ability to conduct evaluations of library program. This guide will only skim the surface of the topic of evaluation; it is by no means comprehensive. At the end of the guide there is a list of resources on specific subjects to consult for more in-depth information. Throughout the guide, certain evaluation terms are stylized in **bold text**. These terms are described more thoroughly in the glossary at the end of the guide.

Privacy and Library Evaluation

Protecting user privacy and confidentiality has long been an integral part of the mission of libraries. An ethical approach, including protecting patron privacy is critical to evaluation. In many evaluations there is no need to collect data which personally identifies any individual. Collecting **personally identifiable information** such as names, emails and other data can — but doesn't have to — violate a patron's privacy.

To protecting patron privacy while conducting evaluations:
- Allow individuals to opt-in to any data collection.
- Do not collect data that is "nice to know" but not necessary.
- As much as possible, collect anonymous or aggregated data not tied to individuals' personal data.
- Have a written data management plan for what information will be gathered, where and how it will be stored, who will have access to that data and when it will be erased.

For more information on privacy, ethics and evaluation, see the section on page 9.

Evaluation Steps

Evaluation is a process with structured steps. Outlined to the right are eight steps (plus a "Step Zero") to build a "typical" evaluation plan. Evaluations vary dramatically, but each will go through these steps.

Getting into the Evaluation Mindset

In addition to being a process, evaluation is also a mindset. Learning how to think evaluatively is even more beneficial than learning the concrete steps for conducting a "typical" formal evaluation.

An evaluation mindset is like wayfinding – using the landscape, guides, maps and data to put your whole journey in context and see if you are on the course that is right for you. In practice, an evaluation mindset is about being strategically curious, thinking critically, reflecting regularly and examining underlying beliefs and assumptions. It requires both humility and confidence about your work. Learning to think evaluatively will support your work. Not only can you then conduct evaluations, but you will be more prepared to lead, grow and evolve your programming and to make clear goals and achieve them.

Create a logic model

Step 1: Plan your evaluation

Step 2: Design instruments

Step 3: Collect data

Step 4: Enter and clean the data

Step 5: Analyze the data

Step 6: Interpret the results

Step 7: Create the report

Step 8: Reflect and communicate

Creating a Logic Model ("Step Zero")

A **logic model** is simply a holistic, graphic summary of a program. It identifies the outcomes the program hopes to have and the path the program will take to achieve those outcomes. A logic model is a living document – the model should change as the project changes.

Creating a logic model is technically a step in proper program[1] planning rather than evaluation. This step has been included here because creating a strong design and plan for your program prior to beginning will save you considerable time and stress later on. This guide uses logic models to guide program planning, though other tools such as a theory of change can be used to ground program planning.

Building a logic model will drastically improve the quality of your evaluation and program by:
- Providing a common language within your team.
- Making program assumptions explicit.
- Promoting communication, clarity and transparency.
- Supporting continuous improvements.
- Ensuring alignment of goals, activities and outcomes.

---
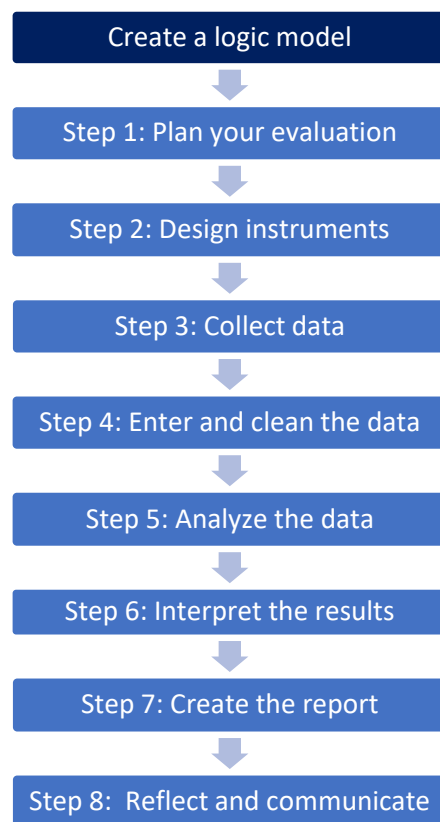
[1] For ease of use, this guide calls all library efforts "programs." Whether your effort is an initiative, program, activity or project, it will benefit from this guidance.

What intended outcomes does this program have?

Ask yourself or your team this question before beginning evaluation, or ideally before launching any program. Begin with the end of the program in mind, using what is known as "**backward design**" to build stronger and more effective programs. In backward design, libraries identify the results they are hoping to achieve first, prior to considering how to measure those results or even what activities might help to achieve those outcomes.[2]

How to create great outcomes and indicators

Evaluators describe **outcomes** as the change we anticipate due to the project's efforts. Outcomes can be intellectual, such as a increased understanding of algebra, or emotional, like increased enjoyment of mathematics.

Outcomes don't have to be at the individual level; projects can have outcomes at the community level, environmental level or institutional level. Outcomes should support your library's strategic plan, and potentially even a city or county plan. Libraries might have internal outcomes, such as change in the system as a whole, or within a specific branch or unit.

How are outcomes different from goals?

Project goals tend to be broad and abstract – with high aspirations such as "appreciation for science" or "increase interest in technology." Goals are useful in explaining a program and creating buy-in. That said, measuring a program's impact from goals alone is challenging, due to their abstract quality. Creating outcomes gives a library specific, tangible aspects to measure that represent the achievement of those goals.

How are outcomes different from outputs?

Outputs are quantified activities, such as the number of kits checked out or the number of e-books downloaded. Outputs don't tend to examine impact in depth. For example, just because a certain number of insect identification kits were checked out over a specific span of time does not demonstrate whether patrons' interest in or knowledge of insects was increased.

Outcomes, on the other hand, should:
- Represent the results or impacts that occur because of program activities and services.
- Be within the scope of the program's control or sphere of reasonable influence, as well as within the timeframe available.
- Be generally accepted as valid by various stakeholders of the program.
- Be phrased in terms of change.

Moving towards expressing the outcomes your program intends to achieve helps tell the story of your program, and allows you to better advocate for your programs.

Now that you have a general understanding of a logic model, you can try building one. On page 17, you will find a worksheet with each step in the building of a logic model described.

---

[2] Different nonprofits, universities, libraries and others refer to the results of programs by different terms. This guide refers to these results as "outcomes."

Step 1: Planning for Evaluation

Now that you have your logic model, the next step is planning the evaluation. Use the Evaluation Planning Worksheet included on pages 19-21 to get organized. While designing your plan, you may need to jot down a few thoughts, get more information, reconsider and then change your plan. Designing an evaluation is as much an art as it is a science. There's not one right or obvious way to construct an evaluation, though there are better and worse ways.

This is not the moment to specify every question you plan to ask your participants, but instead to develop an overall plan for how your evaluation will work.

Evaluation Stages

Specific types of evaluations are useful at different stages in the program. Each of these types has different questions it answers, and possibly different methodologies. You can use all of these types of evaluation to examine your program, perhaps at different times.

Evaluation stage	When you use it	What questions it might answer
Front-End Evaluation *including needs assessments*	• When you are considering creating a program. • When you are trying to understand your users or your community.	• What does the audience need? • What concerns do they have? • What are they excited about? • What program is best for this audience and these outcomes? • What is the audience's incoming knowledge and understanding?
Formative Evaluation *including usability testing*	• At the start of your program or while implementing it.	• Will this work? • Is it working? • How might we improve it? • Is it meeting our audiences' needs?
Summative *or* **Impact Evaluation**	• After your program is implemented and running well enough to determine impact.	• What impact did our program have with stakeholders? • Did our efforts achieve what we hoped to achieve? • Did we have our desired outcomes?

These stages do not have hard and fast beginnings and endings. You may find yourself gathering data about needs at the same time you are trying to improve your program.

The key point to remember is that when you are conducting formative evaluation, any data that helps you build a better program is useful. If you are running an online course, and you are only able to interview five people on how well the course is going, you can still use that information even though it is from a small sample of people.

When you move to summative evaluation, that aspect changes. You need to have a more rigorous and specific plan so that you are collecting reliable, unbiased data to measure outcomes and demonstrate impact. Sample sizes matter in summative evaluation, along with how you recruit, and how you analyze your data. Generally speaking, you'll need larger sample sizes for summative evaluation.

Evaluation Planning Example

Your evaluation will be significantly better if you align the outcome with the method you want to see. Below is just one example.

Sheila E's library is running an afterschool program for middle school girls. It runs for eight weeks. Her team has developed three anticipated outcomes for this program:
- Increased confidence in trying engineering and mathematics.
- Increased interest in engineering and mathematics learning/activities.
- Increased understanding of the STEM educational pathways that can lead to engineering careers.

Sheila E would probably like to get a mix of **qualitative and quantitative data** to ensure the data is reliable and shows a rich picture of change. This will help her get funded again and demonstrate that her library is supporting STEM learning.

She could give the girls a survey when they start and end the program (a pre- and a post-test), but not everyone shows up for each program. She decides she'll do a questionnaire at the end of the program. To make sure they fill it out, she'll give them a certificate for completing the program if they fill out the questionnaire. To measure change, she'll ask them questions like, "Did your interest in engineering change due to this program?"

While that will give her some data, past programs have shown her that she won't get the details that would make her evaluation more compelling.

So, she decides to add a few other elements. She'll ask the students to rate their interest on a poster board at the end of every session. She'll also ask them to sit with the teen librarian when they come in for their certificate and talk about their experience. The teen librarian will do a semi-structured interview, meaning a different representative from the library will ask them a set of questions, but she'll adapt her questions as she goes, tailored to their responses. Finally, she'll check to see whether any of them have checked out any engineering kits, which Sheila E's team selected as a post-program activity.

Having made these decisions, Sheila E can start to fill out her Evaluation Planning Worksheet.

Below is an example Evaluation Planning Worksheet Sheila E might create:

Outcome	Indicator	Method	Sampling Strategy/Sample Size
Increased interest in engineering careers.	Self-rates increased interest.	Survey	**Population**, done at the end of the last session
Increased interest in engineering careers.	Self-rates interest in considering engineering careers.	Interview	Minimum of half the participants
Increased interest in engineering careers.	Discusses career interest enthusiastically.	Interview	Minimum of half the participants
Understanding of the steps needed to become an engineer.	Can describe the steps towards an engineering degree.	Interview	Minimum of half the participants
More confidence in ability to be an engineer.	Self-rates more confidence.	Survey	Population
More confidence in ability to be an engineer.	At least 75% of participants in each session feel more confident about engineering.	Talk-back board at the end of each session	Population

Now that Sheila E knows she is going to do a survey, she can start to think about how she wants to design it. One of the best ways to do that is to use a survey someone else has designed, tested and proven reliable. In this case, Sheila E is focused on STEM for middle school girls, so she'll look at the PEAR website[3] for evaluation instruments. She'll especially be looking for instruments that focus on engineering, career pathways and confidence (see page 10 for more information about designing instruments). One particularly useful tool is the Dimensions of Success Toolkit[4] framework that defines key aspects of a quality STEM learning experience.

In thinking about your own evaluation plan, it is useful to consider the plan from all angles using the following questions:

- How difficult will it be to get your target audience to respond? Will you need to build evaluation into the program itself? Do you need incentives for doing the evaluation?

---

[3] PEAR (Partnerships in Education and Resilience) provides trainings and assessment tools that inform educators about program quality, youth social-emotional development and academic engagement. https://www.pearinc.org.
[4] https://www.pearinc.org/dimensions-of-success.

- Is there a way to make the study a combination of multiple qualitative and quantitative methods? Multiple sources and types of data are always better than one, as they shed a different light and help to cancel out biases a single method might have.

- How will you select a sample and methods? These two items are so tied together it is difficult to consider them separately. Is your target audience online only? In-person? Likely to respond via text message? These factors will change both how you might evaluate and what sort of response rate your evaluation might have. For more on methods, see the Methodology Ideas sheet on page 22.

- Will you need "pre-" program measurements? Or just "post-"?

- Will you have a chance to pilot and then revise your instruments?

- How will you handle data? Will it be entered into a spreadsheet? If so, who will do that?

- When do you need to have the evaluation findings? How much time will you need for analysis? It's easy to underestimate the time it will take to clean and analyze the data, and to write up the findings. The evaluator rule of thumb is that every hour spent gathering data requires two to three hours in the office afterwards to analyze and produce findings. There's a timeline chart (page 20) within the Evaluation Planning worksheet to help lay out when to schedule these things.

- Who will read your evaluation findings? Other librarians? The mayor? Grant funders? Each stakeholder needs different types of information.

Understanding Why Sampling Matters

The **sample** of the data you collect should be a representative subset of the larger group or population. The **population** is the whole group you are working with – all the kids in the day camp, all the families who use the kits, etc.

You should determine whether you need a **representative sample**, which is a sample that approximates the true population. If you are testing out a new website, you may only need 10-15 people to help you. If you are running a summative evaluation, you will need a representative sample.

Representative samples can help minimize sampling bias and error. For example, consider a library that is collecting information on how to serve the community need for a mobile library truck. The library's community is 47% Hispanic, but only 5% of the people it interviewed are Hispanic. This survey data is biased because it does not have a representative sample. In the future, the library could work with a community partner to get better representation and make the sample less biased.

Random sampling occurs when each member of the population has an equal chance of selection. Random sampling is not always possible. For example, you can randomly decide who to survey at your library, but you cannot randomly assign a family to come to storyime.

Population sampling occurs when everyone who came into contact with the program responds. Survey reports should always mention the response rate.

Sample Size

Determining an appropriate sample size is more complex than it seems. The sample size depends on the size and variability of the entire population (the larger group you are trying to generalize about), the smallest sub-group within the population and how much sampling error you wish your study to have. Purely **qualitative** studies can have a smaller number of participants than **quantitative** or **mixed-method** studies. In informal learning settings such as libraries, a sample of 50-80 participants is often large enough for some types of quantitative analysis.

The size of sample does determine its reliability. For example, consider a library that ran two programs: one was a workshop for 30 people and the other was an online presentation that ran several times and had over 100,000 students attend. For the first program, the library has a sample of 29 surveys, which is reliable since over 96% of the people in the workshop responded (the response rate is 96%). If the library were to survey the same group again, their responses would likely be mostly the same, even if only 28 people responded. In contrast, for the online program, if the library received 150 survey responses it would have a response rate of less than .01%. The data in this case is unreliable. The responses might be only from the students who loved the program, or from a certain type of student.

Privacy and Human Subjects Protection

Libraries are known for their protection of privacy for their patrons. At times, evaluation gathers data that may be considered private, and many libraries seek out how to address this discrepancy.

Within the U.S., both evaluation and research are held responsible with protecting the rights and welfare of all people involved in research. These laws are in place to protect the human subjects of research.

Institutional Review Boards (IRBs) were developed to review and approve research with human subjects. Although initially developed for protection of people involved in medical research, obtaining IRB approval is now required for any institution that receives federal funding to conduct research (or evaluation) with human participants.

Many IRBs require researchers and program directors to be certified in Human Subjects Protections in order to submit for review. There are multiple online courses to become certified, the most popular is the CITI Human Subjects Course.

Although most evaluation does not risk harm to participants in the traditional sense, audio- or video-recording people – or collecting personally identifiable information — can invade their privacy in ways that can make them feel uncomfortable. If working with children, other special considerations must be met for protecting their privacy since as minors they are considered a vulnerable population and only parents or guardians can decide on their participation.

Behavioral and survey research also fall under the category of research involving human subjects, so before any data collection begins, determine whether you need to obtain IRB approval for your evaluation – even if your organization does not accept federal funding.

You many need to begin your study with the signature of an informed consent form. This form describes the purpose and procedures of the study, risks and benefits, the voluntary nature of participation, how information will remain confidential and researcher contact information. An example of such a form is within this guide on page 26.

Step 2: Design Instruments

Instruments are the tools used to collect data, such as a survey, an interview form, a feedback board or others. Surveys are one of the most common formats for collecting data. Included at the end of this guide is a checklist for developing better surveys (page 23).

There are many books written about how to design instruments. Yours can be better by keeping the following general principles in mind:

- In the ultimate sense, we evaluate because we want to help our community and hold ourselves responsible to our community and funders. Our goal is not to judge people, or to make them feel inadequate. If the program is not working as we hoped, that is on us as program designers, not on our audience.

- Evaluation instruments that feel like an integrated part of the program will have a better response.

- To more efficiently collect and analyze information, use closed-ended questions whenever possible. These can be multiple choice with possible answers already provided or ranking scales. Open-ended questions require more effort to answer, and more effort for you to analyze, so they should be used sparingly. Open-ended questions are best when you don't know what answers you might receive, or when you need detailed responses.

- How you ask about demographics is a signal to your audience about how much you respect them. Including thoughtful answer categories for gender, race, ethnicity, income and other personal questions will demonstrate that your library honors the respondent as an individual.

Sound Like a Statistician

Validity and **reliability** are key concepts in both evaluation and research. A valid instrument has been shown to measure what it is intended to measure, and a reliable instrument will measure accurately across time and in different contexts.

For example, an instrument that measures enjoyment of science isn't a valid measure of whether someone will pursue a career in science.

HG &CO

- Always allow respondents a moment to express what's on their mind. You can do this in the form of a final open-ended question, such as, "Is there anything else you would like us to know?"

- Respect a participant's time and effort. Do not ask them questions that are merely interesting to know. Only ask them questions that are necessary. Always thank them profusely and sincerely.

Step 3: Collect Data

Now we're finally at the data collection step! Don't forget: People are the experts of their own opinion and knowledge. Don't try and educate or correct them. Your job is to listen.

For in-person data collection:

- Genuinely convey that their thoughts matter and you are interested.

- Always be friendly and accommodating and try to be aware of their needs, expectations and agendas. Make the interview/survey as relaxed and comfortable as possible. Find them a place to sit. Smile. It's okay to be conversational, even though you are asking them very specific questions. You don't want to sound like you're reading off a script, even though you do need to use the language of the survey. Often, in-person interviews are one of the best parts of the program experience.

- Listen. Follow up with non-judgmental prompts if possible, such as, "Tell me more about that." And, "That's interesting, say more." When you convey genuine interest, people will tell you more. And often it's only after prompting that you get some of the most valuable insights.

For online data collection:

- Bring up the survey or evaluation element as early as possible in programs. Remind folks multiple times.

- Personal appeals work better than formal, stiff language. People want to help their community and the library. Let them know how this will help you.

- Always thank people sincerely and, if possible, personally.

One of the best things you can do for your evaluation is to have a plan for how you will deal with the data once you have it before beginning data collection. (The other is to have a logic model!) Based on your Evaluation Planning Worksheet, you should already have some sense of how you need to handle the data.

Step 4: Enter and Clean the Data

Data cleaning is rarely discussed, but it is an important step in the evaluation process. By cleaning, we mean examining your data and making sure it is accurate and in the right format. If you've used Google Forms, SurveyMonkey or another form of digitally-based data collection, the data will already be ready to download and clean.

First, check your data for any errors or misspellings. If you are working with a large amount of hand-entered data, it might be useful to spot-check every dozen rows or so against the original copy to ensure the data is free of errors.

However you arrive at this step, make sure to practice good data hygiene. If you are entering the data from paper into a spreadsheet or Word document, make sure to keep your data organized. For example, if you are entering individual paper survey forms, make sure there is a unique number for each survey within your spreadsheet or document. Then, write that number on the survey, indicating it has already been entered.

Often you will need to decide how to handle missing or incomplete data. Some programs decide not to include incomplete data for analysis. Others save all data but may not include nonsense or obscene responses. These can occur more frequently with middle school and high school audiences. Finally, there may be respondents that do not fit within your sampling plan. For example, they were not at the program in question, or the survey was for senior citizens but answered by a teen.

Step 5: Analyze the Data[5]

After the data is entered and cleaned, you are ready to analyze. "Analyze" in this context means to condense and represent the data, and perhaps conduct statistical tests. Most of the time, the data you will be analyzing will need to be put into descriptive statistics like percentages to show categories. If you have demographic data, it is often easiest to start there.

> **Pro Tip**
>
> If your sample is smaller than 30, don't use percentages, as they can exaggerate differences. For example, in a sample of 15, one person can sway a percentage by over 6%.

With qualitative data, you will likely have to code the data before you can analyze it further. This is a process where you create a set of categories to summarize the responses, label the responses with the categories and then tabulate the number of responses in each category. As part of the process of coding, you'll need to create a "code book" – a document where you've described the categories and given examples of what responses fit in each category.

In practice, this means you need to read over the data and get comfortable with it. You may want to make notes in your first pass. Your goal is to generate four to eight categories, though you may have more than that when you begin. You'll need to have examples of your codes and quotes, so be sure to keep clear notes. You may want to bold quotes to use later in your report.

---

[5] For a list of additional resources on analysis, refer to the back of this guide.

Once you have an idea of your categories, take a sample of your data and see if the responses can fit within your categories. You may need to change your categories, combining them or splitting them into separate ideas. More than eight to nine categories is difficult for people to keep within their head. If you have this many categories or more, you'll need to "clump" several categories together. Alternatively, you may be able to use what is called parent-child coding. In this format, the first code is a more general code, the second is a breakdown within a category. Often you'll find yourself with an "other" category. If the "other" category has more responses in it than several of your other categories, it is a sign you are missing something; your categories are not right and you'll need to reanalyze.

Below is an example of a code book for the question, "If you could change one thing about this museum, what would it be?"

Example of Coding Categories

1 = Wayfinding: All references to navigation and signage, highlights tours and requests for better or more available maps.
2 = Interaction: All requests for guides that could accommodate questions or specialized tours, requests for more "hands-on exhibits," requests for more content for specific kinds of exhibits, any references to updating or changing the content and requests for more interactivity for kids.
3 = Amenities: Requests for seating, bathrooms, escalators (often also coded under "Accessibility" if the respondent stated why) and water fountains.
4 = Physical Improvements: All references to lighting, temp control, crowds and overcrowding.
5 = Languages: References for translations, requests for translations into specific languages and requests for clearer text or labels to be written in plainer language.
6 = Accessibility: All requests for more elevators, wheelchair accommodations or escalators; specific mentions for text legibility or bigger fonts for wall labels.
7 = No Answer, But Offered Praise: Respondents were given "a chance to complain," but decided to say something positive or offer praise.
8 = No Answer: Respondent said "nothing" or declined to give a response.

Once you have the categories coded, you can then create descriptive percentages around the, such as: Wayfinding - 44.6%; Interaction - 20.9%; Amenities - 17.2%.

If the categories are very basic, one person may have no trouble doing the coding. The majority of the time, two or more people should be involved in the coding. One person can develop the coding categories. Then, each person can take a section of the responses and attempt to code them. Finally, you all discuss your codes to see if they match. This is a very informal way of looking at inter-rater reliability. You are making the data analysis more reliable by seeing if a different "rater" would code the same data the same way. If not, then your coding needs either more specificity or you need to change the codes you are using.

Step 6: Interpret the Data

We've broken analysis and interpretation into two steps to emphasize how analyzing the data and interpreting it are different. In the analysis stage, you are pulling together the data and summarizing it. You may be running statistical tests or crosstabs. Those summaries are important, but they aren't enough. The next step is to figure out what the data means.

Every good evaluation needs to place the data within context and to explain to the stakeholders what questions the data is answering, and how to make use of that data. For example, knowing 24% of the libraries' visitors are teens is interesting. But, it is not as useful as knowing whether that represents an increase or decrease in teen visitation, or whether this points to the need for a teen librarian. That context is important.

To begin interpretation, it is often useful to organize the data analysis in an outline by outcome, or if you are writing a formative report, by theme or aspect. Within each section of the outline, consider what the analysis shows. Is the program working the way it was intended? What works well and what works not so well? Why? The why is often the most important part of the interpretation. Your data may not say directly why a particular element happened the way it did. It is acceptable to speculate, but you must base that speculation on data and be clear where you are hypothesizing and where you are simply reporting.

Other questions to consider during interpretation include whether your analysis supports or contradicts results of a previous study, how the analysis compares to what staff thought would happen and whether the analysis raises new issues. What do your results imply for your library and other libraries? What directions would you recommend for the future?

Step 7: Create the Report

An evaluation report contains at least three primary sections: one that describes the project, one that describes the evaluation and one that describes the results. There may be many other elements in addition to the substantive report, such as an executive summary, logic model, recommendations for next steps and more functional elements such as a table of contents, attachments of instruments and so forth.

In the lists below, you can see the specific elements that should be within the report.

Essential Report Components

Describe the project:
- Funders
- Partners
- Goals
- Audiences
- Duration
- Activities

Describe the evaluation design:
- Sample
- Methods
- How the data was gathered
- When the data was gathered
- Number of responses
- Response rate, if available
- Any biases that might occur

Describe the results:
- Summarize results by outcome
- Conclusions
- Recommendations
- When the data was gathered
- Any biases that might occur

Some of these report components may only be a sentence or even part of a sentence, but they make the report more comprehensive and ensure the findings are presented in context. Below are hallmarks of a good evaluation report.

Hallmarks of a Good Evaluation Report

- It uses data to tell a story.
- It organizes results by finding, not by survey question. Survey questions are designed to make it easy to respond for participants. Reports should emphasize critical findings. Use your outcomes as your sections.
- It is honest about what you can and can't conclude from the data.
- It has a mix of visual and text elements.
- It contains direct quotes of participants or stakeholders. Nothing grounds a report like hearing the human side.
- It has pictures of the activities. Show off the photos of your programs!
- It is no longer than the attention span of your stakeholders. Don't produce reports no one will read, instead condense the important points in an executive summary.

In this section we've made it sound as if a written report is the natural outcome of an evaluation. In fact, there are many ways to represent your findings. A short video presentation describing the results might be more useful to your stakeholders than a written report.

Always consider who the consumers of the report will be.

Step 8: Reflect and Communicate

This is one of the most frequently neglected stages of evaluation. Evaluations build knowledge, stronger programs and deeper connections to colleagues and workplaces – but only if they are learned from and only if they are shared.

Once you've done all the work in the previous steps, make sure you talk about it with your peers. Make sure stakeholders see it. Sharing a report is helpful, but making personal connections to the relevance is even better. Evaluations support a learning culture – one that benefits from continuous improvement.

Some approaches to making personal connections when communicating findings include:
- Organizing conversations and discussion groups to reflect on the findings and strategize for the future.
- Doing a presentation for your library team.
- Hosting a Q&A with other branches.
- Sending the executive summary to colleagues.

We've even heard of teams putting outcomes in fortune cookies and placing graphs on lunchroom bulletin boards. Don't be afraid to get creative!

Creating a Logic Model Worksheet

There are a couple different places you can start, and no place is wrong. Here are suggestions:

1. Start by writing down all of the **Outcomes**. You might want to divide these into primary and secondary outcomes. If you get stuck, move to the audiences or activities and then come back. Ask yourself — what do people, know, think, feel, believe or do differently due to participating in this program?

2. **Audiences.** Write down all of the audiences, being as specific as you can. Don't use euphemisms for race or income — they only serve to obscure the purpose. For instance, instead of "Underserved K-12 students," write "low-income Black middle school students." Increasing how specifically you focus on your core audiences doesn't rule out impact for others.

3. **Activities.** Fill in your activities next. Activities are processes or actions your program undertakes to achieve the goals, i.e. the stuff your program does or offers. For example, these might include offering classes or training, 3-D printing services, kits to check out, storytimes or business incubation. Try and balance being specific without listing hundreds of items.

4. **Inputs.** Next, write down all of the inputs for your program. Some logic models leave these out, but it can be useful in both explaining your program and reflecting on it to list the inputs. Inputs are the building blocks or ingredients of your program. They might include grant funding, partners in the program, staff and volunteers, advisors, facilities and supplies.

5. **Revise.** After your first pass of the model, review and edit. Show your model to others, both those working on the program and those who are not. Use the guidelines below to help iterate and improve your logic model.
   - Identify gaps and assumptions. Do the different components of your logic model relate to each other logically? You should be able to move through the model in an if-then relationship. For example, if middle school girls receive training in coding, their confidence in coding will increase. This example assumes the training will increase confidence.
   - Get outside perspective. It's a good idea to get feedback from colleagues and stakeholders, especially if they haven't worked on this logic model alongside you.
   - Adequate resources. Does your program have adequate resources to implement the activities and achieve the desired outcomes? If you need further resources or buy-in, is that reflected in your activities?
   - Change-oriented. Have you expressed your outcomes in terms of change? Have you identified who/what will experience that change and over what time period?

See the following page for a blank Logic Model Worksheet you can use, as well as an example of a filled out worksheet on page 18.

Logic Model Worksheet: Project: _____

Goals

Inputs

Audience(s)

Activities

Outputs

Outcomes

Assumptions

Logic Model Worksheet: Project: ___Chicago Public Library___

Goals
- To build community (intergenerational and peer-peer).
- To connect STEM to the lives of local Chicago youth.
- To normalize inclusion and equality in programming.

Audience(s)
- middle school aged youth

External Factors
- Participant retention
- Transportation technology
- Content partners
- Neighborhood violence
- School homework

Activities
- Two series of programs across three libraries, centered around a design-challenge concept, culminating in a city-wide event.
- Tween advisory board meetings from September through May at three branches
- Participation in intergenerational mentorship activities.

Outputs
- Repeat participation and engagement in activities.
- # of repeat participants
  - Advisory board
    - Programs
- # of adult mentors involved
- # of programs

Outcomes
- Reflective articulation of new STEM knowledge, resources, and skills demonstrated through presentation of design challenge products
- Reflective articulation of social/emotional skills for 21st Century learning.
- Increased excitement of and awareness for real-world STEM.
- Awareness and utilization of library resources in STEM programming
- Youth voice is incorporated into the project design

Assumptions
- The concept of 'Chopped' will resonate with this audience.
- This audience will be able to articulate what they have learned in the given time frame.

Appendix A: Evaluation Planning Worksheet
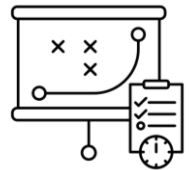
Name of project: _____

Evaluation director: _____

Evaluation team: _____

Stakeholders: _____

Who will be reading the report? _____

When are the findings needed? _____

What will happen or change due to this evaluation? _____

_____

_____

IRB needed: ❑ Yes   ❑ No ❑ Not Sure

Program audience(s): _____

_____

Recruitment strategy: _____

_____

Incentive: ❑ Yes   ❑ No _____

Evaluation Planning Worksheet (Continued)

Evaluation Timeline

Task	Dates	Team member responsible
Identify key stakeholders.		
Hold initial meetings with stakeholders.		
Decide whether you will need IRB approval.		
Develop logic model or theory of change, including: • Develop and prioritize evaluation questions. • Create outcomes. • Determine indicators.		
Develop study design and data collection strategy.		
Develop draft instruments or obtain existing ones.		
Pilot and revise instruments (if applicable).		
Begin data collection.		
Gather and clean data.		
Analyze data.		
Interpret data.		
Create graphics.		
Write/revise report.		
Disseminate report results.		

Evaluation Planning Worksheet (Continued)

Outcome	Indicator	Method	Sampling strategy /sample size	Team member responsible	Dates to collect

HG &CO

Appendix B: Methodology Ideas Sheet

Data collection method	Best used for	Drawbacks and considerations
Questionnaires and surveys	All stages.	Participants can have survey fatigue if they are asked to complete too many.
In-person interviews	Front-end or summative evaluation.	Will provide more in-depth information. Time-consuming for data collection and analysis.
Polls or pulse-checks	All stages.	Can be incorporated into onsite or online programs as rapid-response feedback.
Telephone or online interviews	Front-end or Summative evaluation. Formative evaluation through interviews of internal stakeholders.	Time-consuming for data collection and analysis.
Space observations	Finding out how people use a space through counting, tracking or detailed observations.	Often provides data at the output level without addressing deeper questions.
Participant observations	Understanding how participants react to a program and to one another.	Remember to plan what behavior or conversation you will be looking for and how you will record or note it.
Website analytics	Assisting in understanding overall use and patterns. Best paired with other methods.	Often provides data at the output level without addressing deeper questions.
Comment walls and talk-back boards	Quick responses for formative evaluation.	Not a representative sample. Tends to have disproportionate responses from individuals who are either very pleased or very unhappy.
Portfolio creation	Summative evaluation programs where participants learn a skill, and then have a tangible product they create.	Must be built into the program. Participants generally do not create a portfolio on their own time.

Appendix C: Checklist for Surveys and Questionnaires ✓

☐ Keep the survey as short as possible. Often, the longer it is, the quality of the answers or number of responses goes down.

☐ Be consistent with scales. For example, don't use both four-point scales and seven-point scales within a single survey.

☐ Make sure you are including "Other," "Don't Know" or "Not Applicable" when necessary.

☐ If you feel respondents might not be reading the full question, consider reversing some items to ensure they are slowing down.

☐ Ask one question at a time. Avoid "double-barreled" questions such as, "Do you think our website is interesting and useful?"

☐ Check your questions for bias in wording. Wording makes a considerable difference. "Do you approve of the new tax bill?" vs. "Do you like the new tax?"

☐ Be wary of hypothetical questions. It's difficult to predict from a hypothetical, especially within a survey and without much context.

☐ Use clear, unambiguous terms. Don't use words like "most" or "often" if you can avoid doing so because they're open to interpretation; e.g., "Did you use most of the kit?"

☐ Do not use unfamiliar words, abbreviations or jargon (e.g., MBA, hands-on exhibits, interpretation, STEM).

☐ Use questions and categories that you have used before, especially for demographics.

☐ Put demographics questions at the end (unless you are targeting or screening for specific users).

☐ Pilot your survey!

Appendix D: Program Checklist

Example Public Library   **Program:**
Children's Services and Family Engagement **Date:**
                                               Branch:

Adapted from Dimensions of Success (The PEAR Institute) and Six Strands of Science Learning (National Research Council)

	YES	SOMEWHAT	NO	NOTES
ORGANIZATION				
Materials are prepared ahead of time.				
Space is set up for the activity beforehand.				
Space is appropriate for the activity, with limited distractions.				
Activity is well-organized in terms of structure, time management and room management.				
STEAM CONTENT				
Activity encourages interest in and excitement about STEAM.				
STEAM content is accurately conveyed during the activity.				
Activity is hands-on.				
Activity is "minds-on" (i.e., participants are thinking deeply, making connections and developing their own understanding of STEAM concepts).				
Participants use STEAM practices during the activity, e.g.: asking questions, observing, testing, drawing conclusions, planning and redesigning.				
LIBRARY RESOURCES AND TECHNOLOGY				
Activity is tied to books or other library resources.				
Books and other library resources are available during the activity.				
Books and other library resources are used during the activity.				
INTERACTIONS AND ENGAGEMENT				
Participants are greeted at the beginning of the activity.				
Positive group interaction is promoted through introductions, icebreakers or opportunities for collaboration.				
Participants are collaborating through active conversation about their projects.				

Majority of participants (more than half) are engaged throughout the activity.				
Participants exercise creativity to address the design challenges.				
Participants are given opportunities to connect the activity to their own lives.				
REFLECTION AND PRESENTATION				
Facilitator provides opportunity for participants to reflect on the activity.				
Participants discuss what they have learned during the program.				
Participants effectively communicate about their experience and STEAM learning through a presentation to the group.				

Appendix E: Example of an Informed Consent Form

Climate Lab

Participant Consent Form

Dear STEM Lab Participant:

Thank you so much for agreeing to let your child be interviewed here today at the Example Public Library regarding the development of *Climate Lab*.

The purpose of these interviews is to gain feedback on some new game approaches we are considering for the online version of *Climate Lab*. The session will last approximately 20 minutes and we will ask your child about their favorite games, what they think of our ideas and what suggestions they would make. In order to capture all of the discussion, this session will be audio-recorded. To protect your responses, no names will be used on the interview form. Rather, your answers will be given a code number. The link between your name, email and code numbers will be shredded after the evaluator speaks with you.

There are no costs or risks to you associated with this session. As a compensation for your time, you will be given a $20 Example Store gift card. If you have any questions after the session, contact Jane Doe at Jane@DoeResearch.org.

Sincerely,
Jane Doe
Project Evaluator

If you are willing to participate, please fill in the information below:

> **_____Yes, I agree to be interviewed about potential game approaches for the new design of *Climate Lab*. I understand that I will be asked about my game preferences, thoughts about proposes and game components and whether I would recommend any changes. I consent to having these sessions audio-taped. I understand I am free to decline to participate at any time. I further understand that my answers and audio recordings will be confidential, assigned a code number and available only to the project staff.**

Participant's Name (Print): _____

Parent's/Caregiver's Name (Print):_____

Parent's/Caregiver's Signature: _____ Date: _____

Appendix F: Evaluation Glossary

Activities
What a program does to effect change toward the goals. For example, public events, classes, workshops, presentations, mentoring sessions and exhibits are all activities.

Analyze
To examine methodically and in detail the data and then describe that data through words, graphics and the results of statistical tests. This may include compiling descriptive statistics, running statistical tests and crosstabs.

Backward Design
Backward design is the process of designing programs or other learning experiences to achieve specific goals. Backward design begins with the goals — for example what teens are expected to learn and be able to do — and then proceeds "backward" to create programs that achieve those desired goals. Backward design helps to create programs that are focused on the goal rather than the process or logistics of the program.

Bias
Any element what produces systematic error in a research finding. For example, perhaps girls within a specific middle school program are more likely to underestimate their knowledge.

Co-Creation
Co-creation is a form of program design where the participants or patrons are involved in the design of the program. Co-created programs have participant stakeholders involved in the defining of the need for the program, and creating appropriate activities to service that need.

Coding Data
Coding data is "translating" data from one format into another. The codes are the rules specifying *how* to convert data from one format into another. Often this phrase is used for open-ended responses or interviews, which then need to be coded into a format that can be more easily represented or grouped.

Data
Information collected systematically. Data is often thought of as statistical or quantitative, but may take other forms as well — interviews, observations, oral histories, website analytics, etc.

Evaluation
A form of applied research to determine information — frequently stakeholder needs, implementation effectiveness or program impact.

Focus Group
A qualitative evaluation methodology involving an intensive group discussion. Best used for front-end research.

Formative Evaluation
Evaluation undertaken to find ways to improve or redesign a program in its early stages. Focused more on program processes than on outcomes.

Front-End Evaluation
Evaluation undertaken at the early design stages of a program to gather information necessary to inform program development. This may include needs assessments, stakeholder interviews and other formats.

Human-Centered Design
A problem-solution form of evaluation, which focuses on the needs, contexts, behaviors and emotions of the people the solutions will serve. This type of evaluation happens before the program or activity is decided on but after a problem has been identified. It requires research where a program creator can observe thoughts, actions and experiences of patrons. Human-centered design places a premium on empathy — a deep understanding of how and why people behave as they do.

Impact Evaluation
See Summative Evaluation.

Indicators
Specific, observable and measurable characteristics, actions or responses that demonstrate whether a desired change has happened toward an intended outcome.

Institutional Review Board
An IRB reviews a study to ensure compliance with all federal guidelines and other ethical practices. Most IRBs are at universities, though there are private IRB companies as well.

Instrument
An instrument is a tool used to collect and organize information about project participants either through self-reports or observation. Examples include surveys, questionnaires, tests, checklists and rating scales. When deciding on what to use to measure your outcomes, you have essentially three choices: 1) use an existing instrument that has been shown to be valid and reliable, 2) tailor an existing instrument to your needs or 3) develop your own instrument.

Interpret
Interpreting is the act of examining data and analysis, placing that information in context, and framing the meaning of the data. As an example, you may find your analysis has the result "42." While "42" represents the analysis, it is unhelpful as there is no interpretation to tell you what that analysis actually means.

Logic Model
A logic model is simply a holistic, graphic summary of a program. It identifies the outcomes the program hopes to have and the path the program will take to achieve those outcomes. A logic model is a living document – the model should change as the project changes.

Mixed-Method
A study that uses multiple methods, some qualitative and some quantitative, in order to present a more in-depth and reliable understanding.

Outputs
The volume of a program's actions, such as products created or delivered, number of people served and activities and services carried out.

Outcomes
Expected changes in those served by a program, often expressed in changes in knowledge, skills, attitudes and behavior.

Personally Identifiable Information
As defined in OMB Memorandum M-07-1616, PII refers to information that can be used to distinguish or trace an individual's identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual. PII includes Social Security numbers, mailing or email addresses and phone numbers, but can also include an IP address, login IDs, social media posts, geolocation data, biometric data or digital images.

Population
A group of people, events or other subjects of study one wants to describe or about which one wants to generalize. To generalize about a population, one often studies a sample that is meant to be representative of the population.

Population Sampling Strategy
A strategy for collecting data where one collects from the entire population. For example, data is collected from all people within a particular group, or all events held.

Qualitative
Non-numerical data, including attitudes, experiences and behaviors collected through methods of observations, one-to-one interviews, focus groups and similar methods. Qualitative data emphasizes a depth of understanding – a way of understanding complex phenomena rather than how well the data generalizes to a larger population.

Quantitative
Data which can be categorized into numerical patterns, often used to make comparisons between categories by using statistical tests to establish the relationships between variables in order to generalize to a larger population.

Questionnaire
A group of written questions to which individuals respond.

Reliability
Freedom from measurement or random error. The consistency or stability of a measure from one use to the next.

Response Rate

The percentage representing the proportion of the sample who responded to a questionnaire or other methods. Low response rates are one of the more frequent sources of bias in evaluation. Online surveys in particular tend to have low response rates, often making the data unreliable. (See Reliability.)

Sample

A group of subjects selected from a larger group in hope that studying this smaller group (the sample) will reveal important information, trends or findings about the larger group (the population). (See Population.)

Service Design

An approach focused on quality improvements. Service design involves planning and organizing, people, infrastructure, communication and material components of a service to improve the quality of the interaction between the service provider and its customers.

Stakeholder

Either an individual, group or organization who influences, is involved in or impacted by the outcome of a program. They have an interest in the success of the program and can be within or outside the organization that is sponsoring the program.

Survey

See Questionnaire.

Summative Evaluation

Evaluation conducted in the later stages of a program to assess its impact or determine how well the program has met its goals. Often done for outside funding agencies to determine whether program funding should be continued or expanded.

Theory of Change

A theory of change gives the big picture and summarizes work at a strategic level, while a logic model illustrates a program and implementation-level understanding of the change process. In other words, the logic model is like a microscopic lens that zooms in on a specific implementation within the theory of change.

Validity

A measurement that is freed from systematic error – i.e., it measures what it is "supposed" to measure. For example, if a question asks middle school students whether they enjoyed a science exercise, the results would not be valid for whether they gained skills in that exercise. Validity requires reliability, but the reverse is not true.

Appendix G: Evaluation Resources

On Evaluation in General:
A Short Primer on Innovative Evaluation Reporting by Kylie S. Hutchinson (2017).

Eval.org, website of the American Evaluation Association.

Evaluation: A Systematic Approach by Peter H. Rossi and Mark W. Lipsey (2019).

Practical Evaluation Guide: Tools for Museums and Other Informal Educational Settings by Judy Diamond, Michael Horn and David H. Uttal (2016).

Research Design: Qualitative, Quantitative, and Mixed Methods Approaches by John W. Creswell and J. David Creswell (2017).

On Program Evaluation:
"Outcomes Based Evaluation Basics" by the Institute of Museum and Library Services, https://www.imls.gov/grants/outcome-based-evaluation/basics.

Program Evaluation: Embedding Evaluation into Program Design and Development by Susan P. Giancola (2020).

The Managers Guide to Program Evaluation: Planning, Contracting, and Managing for Useful Results by Paul W. Mattessich (2003).

On Logic Models:
Logic Model Development Guide by the W.K. Kellogg Foundation (2004).

The Institute of Museum and Library Services' Community Catalyst Initiative (CCI), https://www.imls.gov/cci/tools-resources.

The Logic Model Guidebook: Better Strategies for Great Results, Second Edition by Cynthia C. Phillips and Lisa Wyatt Knowlton (2012).

On Survey Development:
Designing Quality Survey Questions by Sheila B. Robinson and Kimberly Firth Leonard (2018).

Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method 4th Edition by Don A. Dillman, Jolene D. Smyth and Leah Melani Christian (2018).

On Privacy/Human Subjects/Institutional Review Boards:
"Librarians and the Institutional Review Board (IRB): Relationships Matter" by Laureen Cantwell and Dorie Van Kampen-Breit. (2015). Collaborative Librarianship: Vol. 7 : Iss. 2 , Article 4. https://digitalcommons.du.edu/collaborativelibrarianship/vol7/iss2/4.