Making Your Program Evaluation Work for You

A Manager's Guide to Evaluating Coastal and Marine Education Programs

CALIFORNIA COASTAL COMMISSION

Making Your Program Evaluation Work for You:

A Manager's Guide to Evaluating Coastal and Marine Education Programs

Table of Contents

(Ctrl+ click on a section to go there directly)

About this Guide....3
Evaluation Basics....4
Stage 1: Preparing for Your Evaluation....8
Stage 2: Designing Your Evaluation....18
Stage 3: Data Collection and Analysis....27
Stage 4: Reporting Results and Using Findings....35
Learn More....36
Appendix A: Logic Model Template....37
Appendix B: Data Collection Methods Chart....38
Appendix C: Detailed Description of Methods Options....40
Appendix D: More Information on Random Sampling....47
References....48

About this Guide

The California Coastal Commission created this guide as a resource for its Whale Tail® grantees and others contemplating or conducting small to medium sized environmental education programs.

The guide is directed to a project coordinator, manager, or grant writer who is not an expert, but has at least a basic understanding of evaluation. It uses the terms “project” and “program” interchangeably, since our grants are awarded for a wide range of activities.

The guide walks the reader step-by-step through the process of planning and conducting an evaluation, and provides links to additional resources to delve further into specific topics. It encourages the reader to invest time into planning their evaluation such that it is tailored to their needs and objectives and yields meaningful information. It encourages thinking broadly about methods and combining quantitative with qualitative methods.

The guide is a compilation of material from a range of evaluation handbooks, websites, and other resources, as listed in the “Learn More” and “Reference” sections, as well as many other places within the guide. A special thank you to the University of Michigan’s MEERA website, and the NOAA evaluation guides; both were consulted extensively and are referenced within.

Written by:
Christiane Parry
California Coastal Commission

Evaluation Basics

Evaluation is a basic activity. We are all constantly evaluating, usually in an informal way, as we make decisions throughout our lives. On a day to day basis we decide what to wear (is it hot or cold outside?) and what to eat (what am I in the mood for, how hungry am I, what's in the refrigerator). We read labels and compare items in a grocery store. Even with the projects and programs we conduct, we are evaluating them informally, whether or not we conduct a formal evaluation.

Program evaluation is a tool that brings this process out into the open and adds structure and intentionality to the process. The results can be used to inform decisions, improve operations and outcomes, guide strategy, and increase understanding.

In other words, at least in our context, evaluation is about providing information to give us insight into what is happening within our programs and support us in doing our jobs better. We do *not* want evaluation to be about creating a paperwork burden for our grantees without adding value. Our goal is to avoid this pitfall and help our grantees design and carry out evaluations that are helpful.

That said, does every program benefit from formal evaluation? Is it always a good use of resources? In our experience, most programs will benefit from some level of evaluation activity. The appropriate level, however, varies considerably and depends on factors such as the program's stage of development, the intended uses of the evaluation, and available resources.

A good evaluation is *tailored* to the program. What do you most need to know? For example, if you are pilot testing a new program, you will want to focus on implementation, and on getting feedback that will help you understand and improve it. This would be a different exercise from evaluating a proven, well-established program that been evaluated in the past. In that case, evaluation might take the form of monitoring – did implementation go as planned, are the outputs and indicators what is expected?

What Purposes Can Evaluation Serve?

Evaluation can serve a wide range of purposes. Often a single evaluation will have multiple purposes. Here's a partial list:

- To support project planning.
- To understand how a project is implemented and/or to facilitate project improvement.
- To assess merit and worth or, more specifically, to determine if a project is meeting its goals and objectives. A related purpose is to determine what the outcome is for participants.
- Accountability to funders or other decision-makers.
- To better understand a project and its context (organizational, community, etc.).
- Public relations.
- To build evaluation capacity and evaluative thinking.

Types of Evaluation

Another way to slice and dice evaluation is to talk about the types of evaluation that one might do. There are many types of evaluation and different terms used to describe similar types. Some types of evaluation describe or prioritize a particular prescribed methodology, or a philosophical approach that defines a set of lenses and priorities.

One common way to categorize types of evaluation at a basic level is based on the stage of a program. *Formative or process evaluation* focuses on the questions that come up during program implementation. The main idea is looking at how a program is being carried out and ways to improve a program. *Summative or outcome evaluation* focuses on the questions that come up around program outcomes and impact. For example, did (or does) the program achieve desired changes in knowledge, attitudes, or behaviors? Were there unexpected outcomes?

For some programs, it may make sense to conduct separate formative and summative evaluations – a complex program may not be "ripe" for summative evaluation for several years, and it might be important to analyze and draw conclusions based on a formative evaluation in the meantime. However, given their project scopes, for the majority of Whale Tail® grant funded programs, the same evaluation can look at both implementation and outcomes.

Another type of evaluation typically takes place before a program is undertaken. Although not the focus of this guide or Whale Tail® grant requirements, *front end or planning evaluation* can be very helpful in guiding a program's development. The basic idea is that gaining and applying an understanding of the context in which a project will or is operating will make for a better project. If you are in a situation where this

type of work is possible, it can help head off problems and obstacles down the road.

Examples of front-end evaluation activities include:

- Assessing the needs, assets and resources of a target audience.
- Conducting a "gap analysis" of existing similar projects in an area to get the lay of the land and define what is most needed.
- Evaluating the external environment, e.g. political, organizational, etc., in which your project will be or is operating – what external factors influence the success of your project, what do you need to consider as you plan your project?
- Evaluating the viability of an innovative project idea. This could involve piloting aspects of your project before the plan is finalized.

Benefits and Uses of Evaluation

In the environmental education world as elsewhere, a lack of time and resources and wanting to focus on "doing" versus measuring, are very good reasons that many excellent education providers place evaluation low on their priority list. Although the landscape is changing, the benefits of evaluation are less well appreciated.

The value derived from evaluation work lies in the extent to which it is useful and used. In this regard, it may be helpful to think about "use" in a very broad sense to include:

- Directly inform program decisions and improvements
- Deepen the understanding of program staff and stakeholders regarding a program
- Indirectly feed into broader decisions and policy that go beyond the program itself

A fourth category is known as "Process Use," which refers to the learning and changes that occur for individuals and at the organizational level, by virtue of involvement in the evaluation.

In developing this guide, we interviewed people from environmental education organizations that are doing strong evaluation work. Each had a similar story – for various reasons (grant requirements, staff interest) they had been prompted to take evaluation to the next level within their organization.

Each stated that this shift had led to tangible benefits for their programs and for the organization. Their evaluation activity had helped them to refine and

improve programs, and in some cases let go of programs that were not working. This work had also created a shift within the organizational culture, prompting staff to think more strategically about what they are doing and why, and feeding into program planning activities.

Some Take Away Messages

The ultimate quality and usefulness of your evaluation is dependent on the time and effort put into planning and design. There is a tendency in the environmental education field for program evaluation to follow a formula, relying primarily on metrics and pre-post surveys for data. Of course there is nothing wrong with this design and it may be the most appropriate choice for some evaluations. The problem is that using a set formula ignores the specifics of the program and skips over some potentially important considerations.

This guide will not answer all your questions. For Whale Tail® Grantees, program staff may be available to consult on specific situations. However, for anyone doing evaluation, there will be an element of plunging in and learning as you go. Our hope is that this guide provides some signposts, while giving you encouragement to step outside your comfort zone and try some new approaches.

Organization of this Guide

The sections that follow will take you stepwise through the process of planning and carrying out your evaluation. Or you can of course pick and choose based on where you need guidance.

Stage 1 – Preparing for Your Evaluation is about defining the goals of your evaluation, and thinking through the questions you want to answer.

Spending time on this will help set you up well for **Stage 2 – Designing Your Evaluation**, which is about selecting data collection methods, developing instruments, and creating a plan for data collection.

Stage 3 – Data Collection and Analysis will help you navigate this stage of the evaluation process, how to collect data and what to do with it once you get it.

Stage 4 – Reporting Results and Using Findings encourages you to be strategic about how results are communicated and used.

Stage 1: Preparing for Your Evaluation

Stage 1 is about getting the pieces in place so that Stage 2 – in which you develop your evaluation design – goes smoothly. Some of these pieces may already be known, in which case it may be more a function of gathering the pieces versus figuring them out. For example, you may already have a logic model for your program, or you may already know your evaluation budget and who will carry out the evaluation. These pieces will form the basis of your evaluation plan.

Step 1 – Begin to Put Together an Evaluation Plan

An early consideration is how you are going to document or organize your evaluation. It may make sense to create a plan that will set forth the framework for your evaluation. Depending on the scope of what you are going to do, your plan can include your evaluation budget, the roles of individuals involved in the evaluation, a program logic model, a time-line, evaluation goals and objectives, evaluation questions, and a data collection and analysis plan. These individual items are discussed in this guide.

What Is Your Evaluation Budget?

Although there's a certain amount of guesswork in setting an evaluation budget (as with most budgets), a good rule of thumb is to budget X to XX percent of the program budget on evaluation activities. The cost of your evaluation will vary depending on your evaluation objectives, the plan, and who carries it out. For example, certain data collection techniques and analyses are more time consuming than others. So consider your expectations regarding your evaluation results when setting the budget and cost it out the best you can.

Who Will Carry Out the Evaluation?

Another consideration is who is going to carry out the evaluation, and who else will be involved at some level.

❖ External versus internal evaluator?

Most Whale Tail® grantees conduct evaluations in house, while some use external evaluators. The choice may be a function of organizational culture – "this is the way we do things here." Or it may hinge on resources – hiring an

outside evaluator is generally more costly than using internal resources. It's worth asking the question, however. You also may consider a hybrid arrangement in which the program coordinator takes on some of the evaluation tasks in consultation with an external evaluator. Take a look at the box for some of the pros and cons of hiring out.

Advantages of Using an External Evaluator	Disadvantages of Using an External Evaluator
• **Expertise** – may have evaluation expertise not available on staff • **Objectivity** - concerned with the process, versus particular people or the organization. May also be perceived by others as more objective. • **Outsider perspective** – may see things those close to the program miss.	• **Cost.** Can be more expensive (although possibly offset by efficiencies relative to time spent by in-house staff). • **Lack of familiarity** with the context of the evaluation – may spend time getting up to speed. • **Outsider.** May need to build trust of participants.

❖ Involve other stakeholders?

In addition to the person(s) who will carry out the evaluation, there are other roles that should be identified in advance. If the evaluator is different from the program coordinator, the program coordinator also needs to be involved in the evaluation and should work in tandem with the evaluator. In addition, most programs have a group of stakeholders of some kind. For example, for a school program, you may have multiple stakeholders including teachers, students, administrators, and parents.

If and how you involve these stakeholders will depend on the nature of your program. How invested are the stakeholders in the program? Do they have a vested interest in how the evaluation goes? Do they have expertise that would be helpful to tap into in planning and/or carrying out the evaluation? Depending on your answers to these questions, you may consider adding key stakeholders to your evaluation team in some way e.g. an advisory group that your evaluator could consult with periodically. You can also decide on the desired level of stakeholder involvement, which can range from hands-on involvement in determining an evaluation plan to periodic updates on what is happening.

Planning for Use

It is worthwhile to incorporate the notion of use into all aspects of evaluation planning and implementation. To keep you on track with respect to maximizing evaluation use, think upfront about and keep pondering the following questions:

- Who are the intended users of the evaluation?
- How should they be involved in the evaluation process?
- What are the intended uses of the evaluation results?
- Are there barriers to use of the evaluation? How can they be addressed?
- How can the evaluation results be conveyed in a timeframe and format to be most useful?

For more information:
Utilization-focused Evaluation Checklist (Michael Quinn Patton, 2002.)

Conducting an Evaluation that is Culturally Responsive

Many of the programs funded by Whale Tail® grants involve ethnically diverse, multi-cultural audiences. Hopefully, these programs have been designed and planned in such a way as to be culturally appropriate for those audiences. Similarly, when conducting an evaluation, it is important to take culture into consideration, both to understand the program context and to adapt the evaluation to it. An evaluation that is culturally responsive seeks to understand how the cultural background of participants may influence how they view and respond to the different phases of the evaluation. It also requires making adjustments to ensure that there is not a disconnect between the evaluation and participants. For example, it is important to understand how survey or interview questions might be interpreted by respondents from different cultural backgrounds, and what questions are most salient for the audience.

Similarly, an interviewer who understands the cultural context of an interviewee will conduct a richer, more fruitful interview. The interviewer will be better able to connect with the interviewee, interpret non-verbal cues, and know when and how to probe a response or comment. Ideally, you have staff or stakeholders who share the culture of the target audience inform the evaluation, and help interpret the evaluation through a cultural lens. If this is not possible, developing an understanding of and responding to the cultural context through a careful consideration of this issue during the evaluation process is key.

The following resources provide additional information on strategies for conducting culturally responsive evaluation.

NSF User-Friendly Handbook for Project Evaluation, page 75 Chapter 7, A Guide to Conducting Culturally Responsive Evaluations.
http://www.westat.com/Westat/pdf/news/UFHB.pdf

Ways to make your evaluations more culturally sensitive. Prepared by Ellen Taylor-Powell, Evaluation Specialist, from Preskill, H., & Russ-Eft, D. (2005). (This is a one-page list.)
http://www.uwex.edu/ces/4h/evaluation/documents/Waystomakeyourevaluationsmoreculturallysensitive.pdf

Step 2 – Consider Developing a Program Logic Model (Possibly with Stakeholders)

Many programs will go into the evaluation stage already having developed a "logic model." For those who don't yet have a logic model and are unfamiliar with the term, it is a *graphic representation of a program, and how it is designed to work, and what it is designed to accomplish.* It shows the link between what goes into a program (such as staff, funds, partners), outputs (such as workshops, products), and outcomes (such as initial reactions, new knowledge, attitude change, values shift, increased engagement, behavior change, environmental improvements).

Benefits of a Logic Model

Developing a logic model can help uncover your assumptions and theories about what activities are going to lead to what outcomes and why. For this reason, it is most helpful to develop a logic model in conjunction with program development, but if one hasn't yet been developed it is not too late - it can also be very helpful in guiding evaluation.

Developing a logic model has many benefits and we strongly encourage it. Benefits include:

- It helps to facilitate a shared understanding of a program and what you hope to achieve. It can be a great exercise for getting a program team on the same page.
- It helps to clarify goals and objectives and how you are going to measure them.
- It is also a tool for prompting creative and strategic thinking about your program.

- The process of developing the model may uncover inconsistencies in your program – different ideas about program goals, or unrealistic expectations about what a program may achieve.
- Once implicit assumptions are made explicit, they can be addressed and revised.
- The model can also be useful in helping to explain your program to outsiders.
- Having a logic model will help make the rest of the evaluation go more smoothly.

How to Create a Logic Model

There are many formats for logic models, which sometimes use different terminology to map out the same process. For example, some models use inputs, activities, outputs and outcomes (initial, intermediate, and longer term). Some add external factors; some include indicators or performance measure. Some models insert "Need" before "Inputs." Use a format that makes sense to you. As you develop your logic model, don't feel that it has to be perfect – the model is designed to be a tool, not a definitive statement, and it can always be revised as your understanding of the process evolves.

The MEERA website has a page on logic models.
http://meera.snre.umich.edu/plan-an-evaluation/step-2-clarify-program-logic

It includes resources that walk you through the process of developing a logic model.

More Logic Model Resources

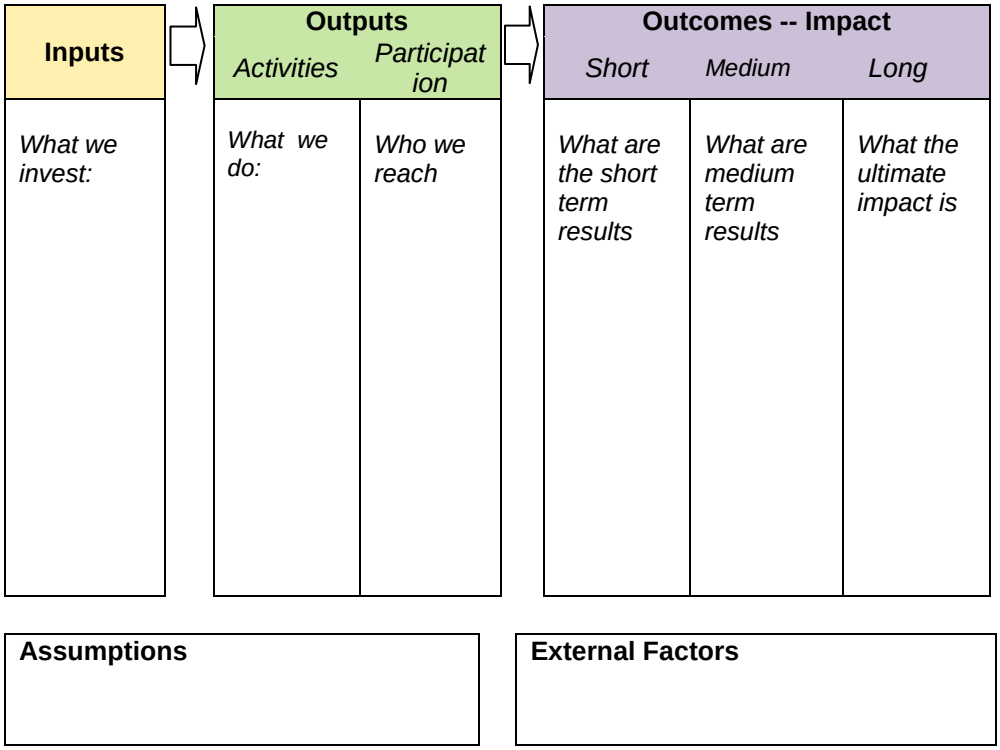
The University of Wisconsin Cooperative Extension has some excellent program development and evaluation resources, including logic model templates that you can customize for your program:

http://www.uwex.edu/ces/pdande/evaluation/evallogicmodelworksheets.html

Point K Learning Center has a "logic model builder" – you will be prompted to enter the activities, outputs, etc. – the tool formats your logic model for you. You will need to register to use this tool, but registration is free.
http://www.innonet.org/index.php?section_id=64&content_id=185

Below is a sample logic model template from the University of Wisconsin Cooperative Extension (see textbox). In addition, **Appendix A** shows a version of this model with examples of the types of items that might fill it in.

These items are written in a generic way. For your logic model, try to be as specific as possible.



As you move from right to left in the model, you become farther removed from results that will be clearly attributable to the program versus other influences in the participants' lives. This is one of the things that makes studying the impact of programs on people so difficult – it can be very hard to isolate the effect of the program. With this in mind, think carefully about your outcomes – what is *realistically attributable to your program's impact*? Or if there are multiple influences and external factors, try to include these in the model, so it shows that the program is part of a larger system of impacts. In addition, in creating your logic model, be clear-sighted in thinking about what changes you might realistically expect to see as a result of your intervention. And keep the program "dosage" in perspective – if it's a low dosage e.g. a one day program, what kind of impact will follow from that – what does success look like?

Step 3 – Establish Goals and Objectives *for the Evaluation*

The next step is to decide on the scope of the evaluation. This is different from the goals and objectives of a program, which are included in the logic model development. The evaluation's goals and objectives will reflect its

purpose, **what** is going to be evaluated, and **how** the evaluation is going to be used. It may be useful to consult with stakeholders on this.

For Whale Tail® grantees, the purpose will in part be accountability – to show how the program has been carried out, and to demonstrate program outcomes. Program improvement is another important objective – evaluation can provide important feedback on what is working, what is not, and why, that can guide mid-course corrections and planning for future iterations of a program.

For a complex program, you may need to focus your evaluation goals and objectives on the specific elements that you most want to get feedback about.

Step 4 – Create Evaluation Questions

Each program will have its own set of questions related to your program goals and objectives. The specific questions you identify will depend on the stage of your program and what you can realistically assess at this point. Your logic model can be helpful in identifying key questions that you would like to explore. Conversely, the process of coming up with key evaluation questions may cause you to revise your logic model. If you have stakeholders engaged in your evaluation process, this is a good topic to involve them in.

Be thoughtful in selecting questions. A good rule of thumb is to ask yourself what difference it would make if you got data to respond to the question. Try to stick to questions that are meaningful and useful – that will have real bearing on your project, that you don't already have the answer to, and that are realistic to gather data to address.

In addition, it is important to craft your questions in such a way that it is possible to collect data and find answers. For example "is the program working well" is not specific enough – "working well" needs to be further defined. A refinement might be "is the program meeting its objectives?" (which are spelled out). Keep in mind as you develop your questions that the questions themselves may indicate the type of data needed. Open-ended and "why" questions may point to qualitative data collection techniques, for example (types of data are discussed in the data collection design section).

Another consideration is whether the question can be answered with the available resources – both time, financial, and expertise. Some questions will require intensive data collection and/or complex analysis – do you have the time, funds and expertise to allow for this?

Examples of Evaluation Questions

Here are some examples of the types of questions an environmental education program evaluation may seek to answer. The questions are set up in a hierarchy from program implementation to results (source: Bennett and Rockwell, 1995). Evidence of outcomes is stronger as you go up the hierarchy, as is difficulty and cost of obtaining evidence. Note: be selective and realistic in identifying questions for your evaluation. We would all like to know as much as possible about how our programs are working, but piling on too many questions, or questions that cannot be answered, is a recipe for evaluation failure.

About program resources:

- What staff and volunteer time was expended, what other resources were used?
- Are financial and staff resources (and any other inputs) sufficient and allocated appropriately?

About program activities and participation:

- The basic facts of what happened. How many participants were there? Who participated – what are the demographics of participants? How many of your products were distributed?
- To what extent was the program implemented according to plan? What changed and why?

About reactions of participants:

- What is the degree of interest of participants?
- What kind of feedback did you get from participants about how they felt about the program?

About other program outcomes (learning, actions, impacts):

- What did participants learn/gain/accomplish?
- Did (and to what extent did) the participants experience a shift in values?
- Did (and to what extent did) participants experience a change in attitudes?
- Did participants feel more "connected to nature" as a result of participation?
- Who benefited and how?
- What do people do differently as a result of the program?
- How well does the program respond to the initiating need?
- To what extent do the different program elements achieve anticipated outcomes? What factors were most important in determining success?
- Why were aspects of the program successful or unsuccessful?

Step 5 - Select Indicators

Indicators are the things you will measure in order to answer some of your evaluation questions, especially those that are quantifiable, and gauge progress towards certain desired outcomes. Looking at your logic model, ask yourself how you will know the answer to a question, and whether a particular short-term, intermediate, or longer-term outcome is achieved? This is the time to figure out what specific, measurable changes would be indicative of meaningful progress in meeting outcomes. You may have more than one line of evidence towards a particular outcome.

Take for example a hypothetical aquarium exhibit on climate change. One of the objectives is to motivate visitors to sign up for an energy efficiency home audit. In this case, the evidence could be the percentage of visitors who sign up for the audit. The indicator could be that "20 percent of visitors to the exhibit go on the energy audit website" as measured by having visitors click through from the aquarium website. An additional indicator could be people who sign up at the aquarium for the audit.

Here's a second hypothetical example. A program to take 4th graders to a week-long ocean themed summer camp has the objective of inspiring students to pursue careers in ocean-related fields. The evidence could be the percentage of students who express an interest in these fields at the end of the camp. The indicator could be a 20% increase in students who express an interest in these fields, measured through a survey at the beginning and end of camp.

Note that you may not have an indicator for each evaluation question or objective. Some evaluation questions will have a more nuanced response e.g. how does a participant experience a program, or why is a situation occurring. Qualitative methods, such as interviews and focus groups can help respond to these types of questions (More on this later, under data collection design.)

Step 6 – Do Some Research

Before launching into your evaluation design, it can be helpful to look around at what others in similar situations have done, and what has worked well or not so well. There is quite a bit of literature available on the internet. For example, MEERA has an archive of sample evaluations of environmental education programs http://meera.snre.umich.edu/reports-and-case-studies/browse that might be helpful to browse. Or you can research specific issues or questions, such as what are effective ways to measure

attitude or behavior change? It may also be useful to research evaluation issues and practices.

Stage 2: Designing Your Evaluation

Once you know the questions you would like to explore through the evaluation, you are ready to identify data collection methods and a design. The evaluation design questions at this point are:

- What type of data would be most helpful?
- What tools should I use to collect data?
- From whom and when should I collect data?

Although we are presenting these as sequential questions, they are closely related and need to be considered as a group. The answers to these questions begin with your logic model and your evaluation questions (what is it you want to learn about your program?), and will also be influenced by timing and budgetary questions.

At this point, it is time to begin setting down your evaluation plan in writing (if you haven't yet started this process). Your plan will describe all of the points discussed here: the evaluation questions, the logic model, the data collection design – including methods, sampling, instruments (creation and pilot testing), and the data collection plan (see below). In addition, it may be useful to create a timeline for your evaluation, with specific deadlines for each component of the plan.

Step 7 - Data Collection Design

This is where you decide on what kind of data you wish to collect and plan how you are going to collect it.

What Type of Data Would Be Most Helpful?

The specific data collection methods you select will be informed by what it is you want to learn about your project i.e. your logic model and your evaluation questions, as well as budgetary and timing considerations.

At a very basic level, you can think about data and data collection methods as being either "quantitative" – something you can quantify, or "qualitative" –data that is in a narrative form. The most common form of quantitative method is the survey, which can also include qualitative data in the form of open-ended responses. Quantitative data is also collected via document review and testing; these could also yield qualitative data. The most common forms of qualitative methods are interviews and focus groups. Many evaluators advocate a "mixed methods" approach. One advantage of a mixed method approach is that quantitative and qualitative techniques

can complement one another— e.g. interviews can provide context for and shed light on the "why" reasons for a survey finding.

Quantitative methods are good at measuring changes and differences, and at gauging whether benchmarks are reached. Qualitative methods can help fill in the stories behind the numbers. They can reveal outcomes that are not expected (and therefore not measured), illuminate hard to quantify outcomes, and provide the kind of understanding needed for program improvement. Whereas quantitative methods help reveal results over a range of participants, qualitative methods help reveal meaning and how a program is experienced by participants.

Some forms of data collection are very straightforward – for example, quantifying the number of participants, the pounds of trash picked up, etc. Although they do not require specialized knowledge or instructions and therefore are not emphasized here, these are also important project indicators.

What Methods Should I Use to Collect Data?

Once you have considered the type of data that would be most useful, you are ready to select data collection methods. Common data collection methods for evaluating marine education include surveys, interviews, observation, focus groups, document review, and tests or student work.

There are other methods. For example, NOAA's Designing Evaluation for Education Projects (2006) describes two additional techniques: concept mapping and case studies. Concept mapping helps the evaluator understand how someone thinks about an idea. Case studies provide a rich, in-depth description of a project, or of a participant's experience. Although the situations in which these methods are likely to be used is less common, they have important benefits in certain instances.

Appendix B: Data Collection Methods Chart shows the common data collection options listed in the first paragraph, and a brief discussion of some pros and cons – this chart can be used early in the process of selecting data collection methods, to compare the various options and come up with a rough idea of your plan.

Here are links to detailed descriptions of each method
(from **Appendix C: Detailed Description of Methods Options)**

- **SURVEY**
- **INTERVIEW**
- **FOCUS GROUP**
- **OBSERVATION**
- **DOCUMENT REVIEW**
- **STUDENT ASSIGNMENT OR EXAM**

Additional Resources on Data Collection

Here are some additional resources on data collection methods, keyed to different evaluation skill levels
(from http://meera.snre.umich.edu/plan-an-evaluation/step-4-choose-design-and-tools#instruments).

❖ **User-Friendly Handbook for Mixed Method Evaluations (.pdf)**
Frechtling, J. and Sharp, L. (1997). National Science Foundation.
Level = Intermediate. Chapter 3, Common Qualitative Methods, provides in-depth descriptions. Chapter 3 Appendices provide a sample observation instrument, interview guide, and focus group topic guide.

❖ **Evaluating Environmental Education (.pdf)**
Stokking et al. IUCN (1996). Commission on Education and Communication.
Level = Beginner to Intermediate. Provides guidance on how to construct relevant instruments.

From Whom Should I Collect Data?

The question "from whom" you will collect data relates to your audiences (sometimes there are multiple audiences), and whether or not you will be collecting data from everyone involved or a sample of participants.

If you have a manageable group of participants, sampling will not be necessary – you can collect data from everyone. However, if you are working with a large group, you can consider collecting data from a portion of the population you are working with. Sampling can also make your budget go farther, and therefore may facilitate using more than one data collection method (mixed methods).

There are different types of sampling approaches. In general, a "**random sample**" is preferred – only with a random sample can you make inferences about your population as a whole from the data you've obtained from your

sample. In a random sample, each individual has an equally likely chance of being selected for the sample. **Appendix D: More Information on Random Sampling** provides additional details.

There is also a category of sampling called "**purposive sampling**." In purposive sampling, you have an interest in collecting data from a specific subset of your larger group – e.g. your evaluation questions focus on the impact of your program on a particular socioeconomic group. Purposive sampling is often used for interviewing and focus groups – you have an idea of the characteristics of the people you want to interview and select them on that basis.

A third category of sampling is called "**convenience sampling**." In convenience sampling, respondents either self-select – for example, those who choose to respond to a voluntary internet survey, or are selected by the evaluator for convenience reasons – for example, students at a university are often the subject of the social science research studies conducted there. Although there are limitations in what you can say about a larger group from a convenience sample, sometimes you are restricted by cost or logistical limitations to using it. If that is the case, you need to be very careful about how you interpret and report your results – how might the sample you've obtained be different from the group as a whole? If you don't have a clear answer to this question, you will need to couch your findings as only applying to your sample of participants.

Some Ethical Considerations

When collecting information from and about people, top of mind needs to be treating them with dignity and respect. As you develop your evaluation, think through the points of interaction with participants, and consider questions such as the following: What do your participants need to know to be fully informed? What is the best way to convey this information? Are you promising confidentiality and can you follow through on that? Everyone involved in the evaluation needs to be in the loop regarding what the evaluation is for and how the data is going to be used. In many cases, written disclosure and permission will be required. See MEERA for more information.

When Should I Collect Data – My Design "Set-Up"

Your design set-up refers to the points in time when you will be collecting data. Again it all goes back to your questions and logic model. Are you interested in measuring change? If so, it is often a good idea to have a baseline measure of your participants before the program, so you can

compare the results from after the program. This design is commonly called a "pre and post-test design."

Another way to study the effect of a program on participants is to have a control group that did not participate in the program. This is a more challenging design to pull off, but may work in some instances. Or you may be interested in participant satisfaction after a program, and then follow-up usage e.g. for a teacher workshop. In that instance, you may choose to do a post-program survey to measure satisfaction, and then follow-up in six months to find out if participants applied the knowledge or tools from the program. If you are interested in understanding the experience from the participants' point of view, you may want to collect data while the program is happening – for example, interviewing a sample of participants.

The following resources can help you think about design setups and sampling.

- Types of evaluation designs: http://meera.snre.umich.edu/plan-an-evaluation/related-topics/types-evaluation-designs

- More detail on when and from whom to collect data (see page 43 of this pdf) http://intranet.iucn.org/webfiles/doc/CEC/Public/Electronic/CEC/Books/Evaluating_Environmental_Ed.pdf

Step 8 – Design and Pilot Data Collection Instrument(s)

So, now you have a plan for the data collection methods you are going to use, when you are going to use them, and who you are going to administer them to. The next task is to design your data collection instruments – the actual survey, interview questions, observation guide, etc.

Using Standardized Data Collection Instruments

If one of your objectives is related to a psycho-social type of impact from your program – for example, you want your participants to come away from the program with a greater connection to nature – you should consider using a standardized instrument or scale. The reason for this is that psycho-social impacts such as the notion of "connection to nature" are complex and difficult to measure. In many cases, researchers have studied these constructs and developed ways to operationalize or measure them. These standardized instruments have been "validated" i.e. they have been shown to measure what they intend to measure.

Another advantage is that if a range of programs use the same instruments, it will be possible to aggregate results across programs, and to gauge the impact of environmental education more broadly. However, these instruments also have limitations. They may not be culturally appropriate for your audience, and they may not serve your needs in other ways. But it is worth considering, and seeing what is available. In some cases, the instruments can be adapted to meet the needs of an individual program.

Keep in mind that in some cases, you may need to get permission from the developer before using a standardized instrument. However, since they were developed for the purpose of getting used that shouldn't be an obstacle.

Resources to Help You Find a Standardized Instrument

- Assessment Tools in Informal Science (ATIS) is a searchable database of assessment tools used in informal science education and "citizen science" programs, created by the Program in Education, Afterschool, and Resiliency. It includes a variety of tools, some focused exclusively on content knowledge and competence, others dealing with attitudes and engagement. Although not all of the tools are relevant to environmental education, this database includes the following scales, and provides additional context and information for these scales.
  - *Children's Environmental Attitudes & Social Knowledge Scale (CHEAKS)*
  - *Environmental Values Short Form*
  - *New Ecological Paradigm Scale (NEP-R)*
  - *New Ecological Paradigm Scale for Children (NEP-C)*

  These and other instruments can be downloaded from the website. http://www.pearweb.org/atis/tools/search

- The Environmental Scales website, edited by Coral M. Bruni and P. Wesley Schultz, can be found at: www.conpsychmeasures.com. This website lists 24 scales used to measure different aspects of environmental attitudes and behaviors. The website includes the contents of the scales, the authors, and other background. Many of the scales listed are aimed at an adult audience. Scales in this compendium that are appropriate for children include:

  - *Children's Attitude Toward the Environment (CATE)* has a different format from other scales, designed to be accessible to

children. This 25 item scale includes statements about recycling, animal protection, conservation, nature appreciation, and pollution that begin "Some kids…" The children indicate whether they feel like the kids in that group.

- *Inclusion of Nature in Self* also has an unusual format. It is a graphic tool that can be quickly administered. It can be used with all school aged children and adults, and has wide cultural applicability.

- Other scales not included in the above databases:

  ⇒ *Nature Relatedness Scale*. This 21 item scale is most appropriately used with older children and adults. It was developed by Elizabeth K. Nisbet, John M. Zelenski, and Steven Murphy, and measures the affective, cognitive and experiential aspects of an individual's connection to nature. Here is the scale and scoring information. A short version of this scale is also available.

  ⇒ *The Children and Nature Index*. A 16 item scale that may be used with children aged 8 to 12 or so, developed by Judith Chen-Hsuan Cheng and Martha C. Monroe. This scale can be localized for different types of nature experiences. Here is an article that includes the scale. Here is an example of its use by the Royal Society for the Protection of Birds in the UK, which includes information on implementation.

Creating and Pilot Testing Your Own Data Collection Instrument

Another option is to build your instrument around an existing standardized instrument, or model your instrument on an existing instrument. This method allows you to ask questions that are specific to your program and culturally relevant, but also provides an opportunity for your evaluation to tap into the upfront research that goes into the instrument development.

There are a number of resources on the internet to help you create a customized data collection instrument. Some of these resources are listed below.

Keep in mind that you should only collect data that you will use, and to keep your instruments as succinct as possible to collect the data you need. This

respects participants' time and in the case of a voluntary survey or interview, will help increase your response rate.

It's a good idea to pilot test your data collection tool. In pilot testing, the results are not important, it is about uncovering any glitches – ambiguously worded questions etc. – before you administer the tool for real. Try to find people to pilot test your tool that are similar in age and background to your program participants.

Resources to Help You Create Your Own Data Collection Instrument

Many, many books have been written about how to conduct the various data collection methods, from developing the instrument, to collecting and analyzing data. Below is a small sample of available internet resources on the topics of surveys, interviews, focus groups, and observations. These resources are fairly concise and provide helpful tips to guide you.

Surveys

- **Collecting End of Session Surveys (.pdf)** *University of Wisconsin Extension - Program Development and Evaluation Unit*
- **Questionnaire Design: Asking questions with a purpose (.pdf)**, *University of Wisconsin Extension - Program Development and Evaluation Unit*
- **Tips on Writing Survey Questions (.pdf)** *Harvard Program on Survey Research*

On-line Surveys

If it is feasible to administer your survey via email or social media, you should consider using an internet-based survey service such as Survey Monkey or Survey Gizmo. Although there may be a fee involved (unless you have a small number of survey questions and/or respondents), these services will save you time and effort. Data is automatically recorded electronically and you can run reports, create graphs, and easily download your data to a spreadsheet for additional analysis.

Observations

- **Collecting Evaluation Data: Direct Observation (.pdf)**,*University of Wisconsin Extension - Program Development and Evaluation Unit*
- **Participant Observation as a Data Collection Method** *FQS Forum – Qualitative Social Research*

Focus Groups

- **Quick Tips: Focus Group Interviews (.pdf)**, *University of Wisconsin Extension - Program Development and Evaluation Unit*

- **Tips for Conducting Focus Groups (.pdf).** *InSites*

Interviews

- **Conducting Interviews: Tips for Conducting Program Evaluation (.pdf)**
  Wilder Research and the Minnesota Office of Justice Programs

- **General Guidelines for Conducting Research Interviews** From the *Free Management Library (Adapted from the Field Guide to Consulting and Organizational Development)*

Stage 3: Data Collection and Analysis

You've decided on a design for collecting data and have created your data collection instrument(s). The next stage involves implementing your plan – getting necessary permissions and going out and administering your data collection instruments, followed by organizing and analyzing the data that you get back.

Step 9 – Gather data

Before you go out and actually collect data, you will need to consider whether permissions of any sort are needed, and you will need to make sure your data collectors are trained. Then collect away!

Informing Participants and Getting Permissions

Depending on the setting for and type of data collection, you may need to get permission from the people you want to collect data from. Even if written permission is not required (for example a voluntary survey with no identifying information), it is always a good idea to be upfront about the purpose of the study, the time involved, and how the information will be used.

The following link to the MEERA website provides a good description of the ins and outs of informed consent. Please note that the IRB (Institutional Review Boards) approval will NOT be required for the vast majority of Whale Tail® grantees (unless you work for a university or federal agency that mandates such a review).

http://meera.snre.umich.edu/plan-an-evaluation/related-topics/evaluation-consent-and-institutional-review-board-process#informed

Develop Data Collection Guidelines

Who is going to do the data collection? Whoever this is, they will need guidelines to help ensure consistency and quality in your data collection. Written guidelines should address the details of carrying out the data collection, including:

- Any relevant logistics they need to know
- What they should say to introduce the process and help put participants at ease
- Responses to questions that may come up
- How to address any other contingencies you can anticipate

If you are conducting a survey that people will be taking at home or elsewhere on their own time, e.g. an on-line survey, the instructions should be very complete, should include responses to common questions, and in most cases, a way to get in touch with someone if there are questions or comments.

For some data collection methods, especially qualitative methods such as interviewing and focus groups, consider using a data collector who is experienced with these techniques, or providing training on how to use them effectively.

Pilot testing your data collection instrument(s) (Step 8), will help uncover much of what you need to know to train your data collectors in terms of logistics and the kinds of questions that might come up. The other part of the picture is the human administering the instrument (unless it is an on-line survey). Whether or not they are experienced data collectors, in addition to having written guidelines, it is a good idea to rehearse the process of administering a survey or interview with them. They should be very familiar with the questions prior to administering the instrument, and able to clarify any ambiguities. They should also understand the context for the data collection – why the responses are important, how they will be used.

Step 10 – Analyze your data

You're almost there. You now have a pile (or spreadsheet) of data - questionnaires, interview notes, etc. What do you do with it??? Now is the time to go back to your evaluation questions, to review what you are seeking to learn from the data. This is your touchstone. It doesn't mean that you can't add additional questions that hadn't occurred to you but that come out of your data. Just make sure there is a good reason and that you cover the basic questions as well.

You may decide to go for some expert help at this point (if you haven't already), especially if you have a lot of data or some complicated questions in mind. This next section focuses on the most common types of analysis that are likely to be relevant to a Whale Tail® grantee. It provides an overview of the data analysis process, what you can learn from the data, and the choices to be made in doing the analysis. It will not teach you how to conduct the statistical tests, but will point you in the direction of resources for doing this.

Preparing Your Data for Analysis

If necessary (e.g. your data is not already recorded on a computer), you will need to transfer your written data into an electronic form that you can work with for the analysis, for example a spreadsheet, database or statistical software. It is important to use a logical system and to keep track of how you do it. Each respondent will need a unique ID, for example. See the text box for more information.

Next, plan what kind of analytic technique or statistical test you will use for the different questions on the data collection instrument. You will use different techniques depending on whether you have quantitative or qualitative data.

Recording and Cleaning Data

Developing a coding sheet will help you keep track of how data were recorded. A coding sheet lists questions on the evaluation instrument and relevant data information such as abbreviations you've assigned to variables, and how you record unusual data, so you stay consistent. You should also "clean" your data – proof it for typographical errors etc. Here are some resources to help with these (and other!) tasks.

Enter, Organize and Clean Data
Pell Institute and Pathway to College Network – Evaluation Toolkit

User's Guide to Evaluation for National Service Programs, Data Analysis Chapter (.pdf) *Corporation for National & Community Service*

Analyzing Quantitative Data

It is helpful to first have an understanding of the levels or types of quantitative data you may have. These data classifications will influence the type of statistical analysis used.

Nominal data. Here, numbers are assigned to categories, in order to allow for the analysis. Examples of nominal data are ethnicity, gender, occupation.

Ordinal data. The data are also in categories, which follow an order or ranking. A "Likert scale" is an example of ordinal data – the survey taker chooses between options that are in a particular order such as excellent/good/okay/poor – "excellent" being a higher rating than good, which is higher than okay, etc.

Interval data. These data may also simply be called numerical data, and include such data as a participant's age.

Summarizing and Describing Your Data

The first step is to use descriptive statistics to summarize and describe the data. Descriptive statistical analysis provides different ways of looking at your data so you can see what is there.

Descriptive statistics include:

- **count** (for your metrics – number of people, pounds of trash collected etc.)
- **frequency distribution** (for a multiple choice question – how many people selected each possible response)
- **percent distribution** (the percentage or people who responded in a particular way)
- **"measures of central tendency"** (the measure used will depend on if your data is nominal (mode = most common value), ordinal (median = middle most category), or interval (mean = average value)
- **measures of variability** (e.g., range and standard deviation).
- relationships between two variables - **correlation analysis**
- **change in response from point A to point B for the same person** (without being able to assess if the differences are meaningful or attributable to the program.)
- **difference in responses between different groups** (without being able to assess if the differences are meaningful or attributable to the program.)
- **content analysis** (for open-ended questions)

The following resource helps explain descriptive statistics as they apply to evaluation.

Analyzing Quantitative Data (.pdf)
Taylor-Powell, E. (1996). University of Wisconsin Extension Program, Development and Evaluation Unit

Depending upon your evaluation questions, descriptive analysis may be adequate for your needs. For example, in the list of example evaluation questions from Part 1 or this guide (page 15), questions that describe the program and what happened could be answered with descriptive statistics, including reactions of participants. You can even describe how your participants changed and what they learned.

What you will not be able to do is say with confidence that the changes or differences that you see are (1) statistically significant (over what might be seen by chance), or (2) attributable to your program (versus other influences). Answering these types of questions and others require inferential statistical analysis.

Inferential Statistical Analysis – Changes, Differences, Predictions

Beyond describing your data, you may be interested in making judgments about your data. This is where statistical tests that allow you to make

inferences about your data come in. Do boys respond to your program differently from girls? Are there are differences between groups of participants? Did your program cause a change in your participants from point a to point b? Are observed differences or changes "significant," or possibly due to chance? Inferential statistics are also used to generalize findings from a sample to a larger population.

A separate question concerns whether a statistically significant difference that is measured is "meaningful" or "important." In other words, you may observe a very small difference between a pre and post-test, which may be statistically significant – not due to chance – but not meaningful in the sense of achieving what the program hoped to achieve. When you are conducting inferential statistics, the sample size becomes a key factor – you need an adequate sample size in order to make inferences about the population. Power analysis is a test to see if your sample size is large enough.

The type of statistical test that you will need depends on the question you are trying to answer and the form of your data (nominal, ordinal, interval). There are many, many tests of varying degrees of complexity. Here are some of the basic categories of tests, with examples. This description is intended to give the reader a flavor of the kinds of analysis you may use with quantitative data. Whale Tail® grantees will likely be focused on the third category – testing for significant differences i.e. looking at whether there are changes in your participants after your program, or looking at whether different groups respond differently to your program.

- Exploring relationships among 2 or more variables (different tests depending on form of the data)
- Predicting relationships – regression analysis (two variables), multiple regression analysis.
- Testing for Significant Differences – the type of test varies depending on the form of the data:
    - ⇒ Nominal data: Chi Square
    - ⇒ Ordinal data: Mann-Whitney, Sign test
    - ⇒ Interval data: ANOVA (Analysis of Variance) – can also be used with ordinal data that is normally distributed. Variations on ANOVA, including ANCOVA and MANOVA.

The MEERA page on **analyzing quantitative data** includes several links to help you learn more about quantitative statistical analysis for program evaluation.

Analyzing Qualitative Data

Qualitative data includes interview and focus group responses, observation data, and responses to open-ended questions on surveys. In this section we review techniques for the systematic analysis of qualitative data i.e. going beyond the use of qualitative data as anecdotal or background information, to identifying themes, trends and interrelationships. Although analyzing qualitative data can be time consuming, it can yield rich insights that would not be achieved using other techniques.

Systematic Analysis of Qualitative Data

This type of data analysis is sometimes referred to as "content analysis" and is a way of organizing the data. The process involves reading through the data several times to become very familiar with it. Then, identifying themes and patterns in the data and using that information to develop a "coding scheme" – categories that are used to sort and organize responses. The two steps here are developing the categories and sorting the data according to those categories.

Developing Categories from the Data or Using Sensitizing Concepts

Most frequently, the categories are derived from the data themselves – they emerge upon reading the responses. These may be categories and themes that have been articulated in the responses themselves, or terms that are developed by the evaluator.

Another approach is to use or "sensitizing concepts." Sensitizing concepts are developed from social science literature, theory, or a pre-defined framework related to program goals, objectives, or key questions. The evaluator applies the sensitizing concept to the data to see how the concept is given meaning in that particular setting. These two methods are not mutually exclusive; you may have patterns emerge from the data that reflect known sensitizing concepts.

When reading through the data to develop categories, look for themes and patterns that express the meaning of the data. Themes and patterns are related – you may find a pattern of students feeling a sense of accomplishment when learning a new skill in an outdoor education program. This pattern means that "building self-efficacy" would be a theme of the program being studied.

A key consideration at this stage is that you are looking to find out what the data has to tell you, rather than trying to fit the data to preconceived ideas.

These processes will help meaning emerge from the data, it will begin to tell a story.

Typologies

In some cases it may be useful to further elucidate your data by using it to develop a "typology" – a classification scheme. A typology may be based on an observed pattern, or series of themes, either observed in your data or known. Typologies sort some aspect of a project – activities, participants, outcomes – into categories or "types" that allow for comparison. Oftentimes, there is a continuum of types. An example of a typology developed by the National Museum of Natural History describes museum visitors, according to the following scheme: the Commuter, the Nomad, the Cafeteria Type, and the Very Interested Person (Wolf, 1978).

The typology takes a recognized pattern one step further by describing it in a systematic way. A typology of this nature can deepen understanding, make possible different types of analysis, and can assist in drawing conclusions from a set of data. Again, a key consideration here is that the evaluator look for what the data is telling them, versus imposing their own ideas on the data.

Sorting Qualitative Data

In practice, the activity of developing and sorting into categories is unlikely to be linear – it is an iterative process that may involve rereading notes, and revising categories as the analysis progresses. The process of sorting data involves developing a coding scheme or kind of shorthand that identifies particular categories so they can be applied to the data as a whole.

Once you have developed a coding scheme, you can complete your sorting by hand or use computer software. There are a number of software products available, known as computer assisted qualitative data analysis software (CAQDAS), including freeware available from the Centers for Disease Control. Depending on the amount of qualitative data you have, it may be worthwhile to invest in one of these products. The software will help with locating and grouping coded themes. However, CAQDAS should be viewed as a tool and is not a substitute for a human oversight. And a human being will still need to do the meaning-making part of the analysis.

An additional level of data sorting is to create a cross-classification matrix – in other words relating two categories to each other in a matrix form to see if a new typology emerges. An example is a process/outcome matrix, where program processes or components are crossed with a range of outcomes derived from the data.

Understanding and Explanation

As previously described, the iterative process of categorizing and sorting and data will yield ideas and observations about themes, relationships, and patterns. These themes, relationships and patterns can yield insights into the program that can be articulated in a way that helps to tell its story. This may be adequate for your needs. If, however, you would like to use the data to explore **questions** such as "why" – to explain something about your program such as a cause and effect relationship, an additional level of analysis is needed. There are various theories and techniques for taking qualitative analysis to this level. Please see "Learn More" at the end of this section.

Checking Your Findings

There are a number of techniques for checking on the validity of your findings. For example, sharing your findings with others who know the program as a reality check – does this make sense given what they know? Another technique is to look at your findings from a variety of different perspectives - what are alternative ways of making sense of the data; could the data have been organized differently? You can also examine any outliers – that don't fit your finding - to see if these cases shed light on your analysis. Triangulation is another technique. It involves looking at other data sources to see if the findings from your qualitative analysis are confirmed by or consistent with those results.

The MEERA page on analyzing qualitative data includes several links to help you learn more about qualitative statistical analysis for program evaluation.

Stage 4: Reporting Results and Using Findings

Once you have developed your findings and conclusions based on your data analysis, the next steps are to communicate these results to others, and to decide what to do next – follow up actions to apply the findings and results.

Communicating Evaluation Results

There are a variety of ways to do this. In most cases, some type of written report will be called for, but, there are a range of options in terms of format, level of technicality, and detail. And for evaluations in which there is no third party requiring reporting, it is worthwhile considering the best use of time and most effective way of communicating findings. In combination with or instead of a written report, options include a power point presentation, webpage, brochure, newsletter, or article. Depending on the scope of your evaluation, it may make sense to using one of these formats to provide constituencies with progress updates along the way.

For Whale Tail® grantees, we do require interim and final reports, the elements of which are spelled out in your contract. However, you may consider other ways to communicate results to different stakeholders. The important point is to determine which method will be most effective for a given audience. Effective communication is critical to findings being understood and used. In addition to communicating with stakeholders, consider communicating your findings more broadly, through traditional or on-line media, newsletters, articles, and conferences.

Following Up

Don't assume that because the evaluation is completed, the work is done. Ultimately, the value of the evaluation will rest on the extent to which the results are used. As previously discussed, several types of use are possible, and your evaluation should be designed and timed appropriately to facilitate use. In addition, some uses may occur as the evaluation is happening and information is flowing in – possibly motivating mid-course changes.

The important point here is to build into your schedule time to meet with staff and stakeholders to go over the evaluation findings and decide how to apply them. This is the time to map out a plan for next steps – what are you

going to do or do differently as a result of the evaluation findings, how are changes going to take effect, and who is taking the lead on following up?

Learn More

In addition to the links on specific topics shown throughout the guide, here are some websites with a range of resources on evaluation.

- Tips sheets on a range of evaluation topics from Wilder Research and the Minnesota Department of Human Services. Designed for human service program evaluation, many of the resources will also apply to evaluating environmental education programs:
http://www.evaluatod.org/resources.php?searchVar=tip+sheets

- Website for the Outdoor Research Education and Evaluation Center. The goal of this site is to provide online access to academic resources such as articles, essays, papers, and theses about philosophical, theoretical, research, and evaluation aspects of outdoor education.
http://www.wilderdom.com/research.php

- The "Learning Store" at the University of Wisconsin Extension has a range of evaluation and program development resources. These guides (which have been referenced elsewhere on this website) can be downloaded in pdf form.
http://learningstore.uwex.edu/Program-Development-Evaluation-C234.aspx

- MEERA – My Environmental Evaluation Resource Assistant – from the University of Michigan, is referenced throughout this guide, and offers a wide range of resources and links.
http://meera.snre.umich.edu/

- NOAA National Marine Sanctuaries has an education project evaluation page. Designed for B-WET grantees, most of the information on the site is applicable to environmental education evaluation generally.
http://sanctuaries.noaa.gov/education/evaluation/evaluation.html

Appendix A: Logic Model Template with Types of Possible Content (Your content will be specific to your project)

Inputs	**Outputs**		**Outcomes -- Impact**		
	Activities	*Participation*	*Short*	*Medium*	*Long*
What we invest: ***For example:*** Staff Volunteers Funds Research base Materials Technology Partners --- ***Example Evaluation Questions:*** How much staff time? How much money? Etc. Were the resources sufficient? What is the research base for our program?	*What we do:* ***For example:*** Recruit Participants Public Ed Campaigns Social Media Campaigns School Activities School Presentations Organize Beach Cleanups Field work Field trips Conduct Workshops Create educational resources Train teachers Take kids to the beach Teach science Coordinate citizen science Install exhibits --- ***Example Evaluation Questions:*** Did these tasks get accomplished? Were there changes in the activities?	*Who we reach* ***For example:*** Participants Parents Teachers General Public Community groups Businesses Partners --- ***Example Evaluation Questions:*** Did the intended participants show up? Why or why not? Who heard about it? Were participants satisfied? Did partner roles make sense?	*What are the short term results* ***For example:*** Increase in awareness Increase in knowledge Increase in skills Change in attitude New intentions What else happened that was not anticipated? --- ***Example Evaluation Questions:*** Did participants show an increase or change in _____ over the course of the program?	*What are medium term results* ***For example:*** Application of new knowledge Sustaining of new attitudes Use of new skills or behaviors --- ***Example Evaluation Questions:*** Was the increase in _____sustained over the medium term (definable, perhaps 3 to 6 months) Did participants use their new skills/behaviors after the program?	*What is the ultimate impact* ***For example:*** Long term application of new knowledge Long term sustaining of new attitudes and sensitivities Long term use of new skills and behaviors Improved environment --- ***Example Evaluation Questions:*** Was the increase in _____sustained over the long-term e.g. after one year or more? Did the changes from the project lead to improvements in the environment?

Assumptions

External Factors

Appendix B: Data Collection Methods Chart

Method	Advantages	Disadvantages	When to Use
Survey (includes standardized instruments and scales) Quantitative (number and multiple choice-type responses) and qualitative (open-ended responses)	Not costly. Possibility of collecting large amount of data. Versatile – different types of questions possible. Pre/post administration can uncover changes in understanding, perception, and motivation etc. that take place as a result of or concurrent with, the program. Validated assessments are available for a variety of pscho/social topics.	Self-report can be problematic – potential biases introduced. Lack of depth and explanation. Hard to delve beyond questions, harder to know context for responses.	To capture participant demographics, participant reactions, self-report of knowledge, attitudes, and behavior change (or intentions). When data is needed from a large group. When questions are clearly defined and in depth probing is not needed. When you are trying to generalize from a sample of people (random sampling required) to a larger population.
Interview Mostly used to solicit qualitative information.	Allows for a depth of response – follow up questions and delving – that is not possible in a survey. Nuances can be conveyed by visual cues and voice inflections. Can result in new insights.	More costly and time consuming than surveys. Also has potential for biases – interviewee may want to please interviewer. Volume of information may be large and hard to analyze. Need qualified staff.	When understanding the participant's experience and perceptions is paramount. When trying to understand a complex situation. When answering "why" questions. Can be used to illuminate and help explain survey findings.
Group Interview/ Focus Group Mostly used to solicit qualitative information.	Potentially positive effects of group interaction – you may learn more than with individual interviews, uncover and get a range of perspectives on different issues. Potentially less time consuming than interviews.	Similar to interviews, focus groups are relatively costly and time-consuming. Also has potential biases – added bias of people self-censoring in the presence of peers (or being influenced by them).	As part of a needs assessment process, to get reactions to a program or program element. To gain an understanding of a group experience of a program. When answering "why" questions. To illuminate and help explain findings.
Observation Mostly used to gather qualitative information.	Can view operations of a project as they are actually occurring – implemented as planned? Shows project context. Can adapt to events as they occur. Provides direct information about behavior or individuals or groups. Can view anticipated outcomes.	Can be difficult to interpret behaviors. Biases – presence can influence participant behaviors. Can be expensive and time consuming. Need qualified staff.	To gather accurate information about how a project actually operates, particularly about processes. When it is important to understand a project's context and observe situations as they occur.

Method	Advantages	Disadvantages	When to Use
Tests and Student Work Usually used in conjunction with a rubric to quantify results or quantitative test results	Provides objective information on knowledge or skills at a given point in time. Can help identify a problem or deficiency. Results may be easily scored. Individual results can be compared.	May be oversimplified and superficial. Results can be influenced by attitudes. Language or vocabulary can be an issue. Student work (e.g. journals) may be hard to objectively score. Adults may be concerned about how the test results will be used or resent taking tests	To determine the audience's current state of knowledge or skill, and to determine changes over time.
Document Review Could be quantitative e.g. attendance records, or qualitative e.g. meeting notes.	Yields historical information. Doesn't interrupt project. Few biases about information. Can be inexpensive.	May be Incomplete or inaccurate. Analysis may be time consuming. Need to be quite clear about what looking for, and access may pose challenges.	To gather information on how the project operates, or project context without interrupting the project.

Appendix C: Detailed Description of Methods Options

(Adapted from [http://wateroutreach.uwex.edu/use/documents/NOAAEvalmanualFINAL.pdf)](http://wateroutreach.uwex.edu/use/documents/NOAAEvalmanualFINAL.pdf)

SURVEY

What is it?

Data collection instrument through which individuals respond to printed or oral questions. May be recorded by either respondents or data collector.

How many respondents / participants?

25–1000 (typically). The number of surveys is limited primarily by time and money. Some methods (e.g., email surveys) can be sent to any and all members of the audience who have access to email.

Time issues?

20–45 minutes to complete; two to three months to collect data (for a typical mail survey; much shorter for phone, group or email surveys); one or more months to analyze data (assuming most questions are closed-ended).

Cost issues?

For printed surveys, costs include printing, postage, return postage, follow-up postcards. If surveys are administered in person (either individually or in small groups), costs can diminish dramatically. On-line surveys are least costly.

When to use it?

Surveys allow for the systematic and standardized collection of data that can be generalized to the population (assuming proper sampling procedures were followed). Surveys are appropriate when self-reported data about knowledge, attitudes, skills, and behaviors are desired. Because of their format, surveys can be administered to a large number of people individually (e.g., in person, email, mail) or in groups (e.g., participants in a workshop). In addition, surveys are particularly useful when potential respondents are dispersed geographically. When evaluators have a good idea of the types of responses expected, surveys offer an efficient method of collecting information.

What are some of the benefits?

- May be easiest to quantify, summarize, and report on the data
- Time-effective for use with geographically dispersed or large sample
- Large sample size; data can be generalized to population
- Can be less costly than other methods
- Can provide opportunity for expression without fear of embarrassment (anonymity)
- Can (should) be designed to be relatively bias-free
- Questions from other instruments can be used or modified
- Can gather qualitative and quantitative data
- Respondents can complete at their convenience (unless in person survey)
- Oral and on-line surveys are good for gathering information that requires sequencing
- Easily adaptable to a wide variety of environments

What are some of the limitations?

- May have limited provision for unanticipated responses
- Not adaptable once the survey is distributed
- Requires significant time and high level of expertise to develop valid surveys
- Low return rates for some survey formats (e.g., phone, mail) can skew data
- Can be impersonal (written, self-response format)
- Questions may miss true issues
- Questions and answers can be interpreted differently
- People have been negatively conditioned to the value of surveys
- Language or vocabulary may be an issue
- People tend to want to get the “right” answers (even if the questions is asking for attitudes)
- The survey administrator can influence the respondents
- People may hurry through answers without thinking about them

INTERVIEW

What is it?

A structured conversation (Q and A) between two people either face-to-face or via telephone or email, etc.

How many respondents/participants?

Depends on project. Typically 5–30. Time and cost are dependent on number of people interviewed.

Time issues?

Interviews can last 15 minutes to 2 hours or more, depending on depth. Data analysis is often slow and time consuming. (Data analysis can take weeks, especially if long interviews are transcribed and initial analyses are returned to respondents for a check of accuracy.)

Cost issues?

Project managers must factor in the cost of hiring/training interviewers, transportation costs (if interviewers must travel to meet those being interviewed), and substantial data analysis time.

When to use it?

Interviews are best used when in-depth information or a variety of perspectives about a topic, experience, or service are desired. Often, interviews are selected when the issue(s) is complex. Since broad, open-ended questions can be asked, interviews are appropriate when project evaluators do not feel that they can adequately anticipate types of responses. Interviews should also be used when literacy is an issue.

What are some of the benefits?

- Variety of perspectives can be elicited
- Can be very useful way to build rapport with audience/participants
- Can generate broad and deep data about system or product
- Interviewer can clarify questions and ask for clarification of responses (follow-up questions)
- Interviewer can receive additional information in the form of nonverbal clues
- Questions can be adapted if difficulties arise
- Open-ended questions and a reduced amount of structure allow for new (unplanned for) information to be gathered
- Interviewer can ask for more information than people would want to write in a survey
- Respondents use their own words

What are some of the limitations?

- Bias due to data collector's interest and interpretations
- Time intensive
- Self-reporting of participants may bias data
- Discussion can wander from purpose of interview — results may not be focused
- Unskilled interviewers can make respondents feel self-conscious, and may gather poor data
- Variations occur if there's more than one interviewer
- Open-ended responses can be difficult to organize and analyze
- Difficult to capture everything said unless taping the interview
- Usually used with a small sample
- Replication is difficult

FOCUS GROUP

What is it?

A structured, interactive exchange between an interviewer/facilitator and a small group of people.

How many respondents/participants?

Usually 6–10 participants per focus group. More than one focus group is often helpful, especially if there is more than one audience involved in the project (e.g., parents, teachers, administrators).

Time issues?

Focus groups typically last about 1 to 1 1/2 hours. Data analysis requires transcribing the focus group discussion and pulling trends from it. This process is relatively quick if only one or two focus groups have been conducted.

Cost issues?

Relatively inexpensive unless a focus group facilitator is hired. Other cost considerations include: transportation for participants, room rental, food or other refreshments for participants, honorarium for participants.

When to use it?

Focus groups, like interviews, are best used when a variety of perspectives about a topic, experience, or service are desired. Because focus groups involve several individuals, they are particularly useful when there is reason to believe that peer pressure and/or the social nature of the situation will stimulate creativity and/or encourage discussion of conflicting points of view. Focus groups are best used when the topics are narrow or individuals have a limited amount of information about the topic to share – that is, the discussion is focused. A rule of thumb is that focus groups are best used when any one participant could only talk about the topic for ten minutes.

What are some of the benefits?

- Input can come from wide range of people and perspectives
- Participation may have positive public relations impacts
- Can clarify different points of view
- Can really provide a good indication of the root of a problem

What are some of the limitations?

- May represent special interests
- Participants may use as “gripe session”
- May heighten expectations beyond what can be realistically provided
- One participant may influence attitudes and opinions of others
- Need to transcribe and code information for analysis; hard to quantify
- Cannot capture all information without taping session; not all people are comfortable being taped
- Small sample size

OBSERVATION

What is it?

Data collection based on watching an activity or skill and systematically recording what happens. These observations may be made in person (at the time of the event) or using media (e.g., analysis of a video tape recording). Usually done by an evaluator, however in some cases it may be feasible for the observer to also be a participant e.g. a teacher. Although an outsider may be more objective, a "participant observer" with knowledge of the issues and participants may have deeper insights.

How many respondents/participants?

Typically 5–20. The number of people or objects/events observed depends on the subject being observed (e.g., short, discrete activity such as use of a recycle container vs. complex, lengthy activity such as the ability to facilitate a lesson).

Time issues?

Varies depending on what is being observed. Counting the number of cans recycled during a single event (e.g., community festival) requires little time commitment. Observing group interactions over the life of a project requires more complex data collection (and therefore analysis), and a longer time commitment.

Cost issues?

Varies with the amount of time for observation and analysis

When to use it?

Observation allows evaluators to document actual behavior versus reported behavior. For example, how did participants appear during an activity - were they engaged, detached, confused? What was their affect? How many asked questions? Or an observer can document pro-environmental behavior e.g. how many used a recycling station correctly? How much littering occurred?

What are some of the benefits?

- Little interruption of work flow or group activity (if done properly)
- Generates data about actual behavior, not reported behavior
- Can see project in action
- Can provide good in-depth data
- Data collected in context
- An astute observer can recognize interaction problems not easily described by participants
- Observer can follow action at different points

What are some of the limitations?

- Requires process and content knowledge by observer
- Data can be skewed by observer's biases (and skills); observer can disrupt or alter the situation
- Data are not always quantifiable and may require judgments by the observer
- Typically, limited sample size
- Usually time intensive
- Does not provide a direct indicator of how participants view their actions
- Replication difficult

DOCUMENT REVIEW

What is it?

Existing information in the form of reports, historical data, planning and budget reports, organizational structure charts, workshop evaluations, career development reports. Also includes published research and other evaluation reports.

Time issues?

Varies – depends on the number of documents, their availability, and the amount of information being analyzed.

Cost issues?

Relatively inexpensive, using existing documents and data.

When to use it?

Literature reviews are used primarily in early stages of the development of an evaluation and are particularly useful in front-end evaluation. Existing literature (e.g., theory, research findings, previous evaluation reports) and data (e.g., test scores) can provide a baseline. Other forms of existing data (e.g., budget reports, workshop evaluations) can help paint a picture of the intended audience, their perceptions and their reactions.

What are some of the benefits?

- Can be less time consuming than other methods
- Makes use of already gathered statistical data
- Easier to chart changes over time
- Can provide context for and guide understanding of problems
- Little effort or interruption for participants

What are some of the limitations?

- Data synthesis can be difficult
- May not address specific questions
- Data on the causes of problems may not have been collected
- Reports may be incomplete (may lack meta-data)
- Organizations can be hesitant to share if results reflect poorly on the organization or a project
- Accuracy of reports may not be ascertainable

STUDENT ASSIGNMENT OR EXAM

What is it?

Range of options here. Can be a student journal, student generated art or photographs, an assignment. The important thing to consider if using these types of student work is that you need to figure out how you are going to assess the work, in advance. In other words a rubric of some sort, or coding guide (as with qualitative data). Another option is an exam that assesses knowledge or skill level. An exam can be essay, fill in the blank, true/false, and/or multiple choice formats. *For the purpose of this guide, will assume student work has qualitative data, and tests yield quantitative data, although the reverse could be true in some instances.*

How many respondents/participants?

Depends on the type of assignment or exam, and the amount of time it takes to score or assess them. With a purely quantitative test (multiple choice or true/false), scoring is simplified.

Time issues?

Again, it depends on the type of assignment or test. Qualitative information such as journals or open-ended test questions can be quite time consuming to assess. Data analysis, if closed-ended questions are used, should take a short amount of time.

Cost issues?

Inexpensive.

When to use it?

Student work is used to assess knowledge as well as attitudes and behaviors. Exams are used to assess the knowledge or skills. All of these methods measure a point in time; they cannot predict future or past performance. If administered at intervals, tests can provide an indication of change (e.g., increased understanding) over time (pre/post assessments; longitudinal). Student work can similarly look at change over time if several intervals are used.

What are some of the benefits of using student work in evaluation (assumes qualitative)?

- Can be a non-intrusive expression of the student's experience or learning in a program
- May serve multiple functions – student learning as well as program evaluation
- May be implemented during course of school day.

What are some of the limitations of using student work?

- May be difficult to score or assess - subjectivity
- Difficult to discern influence of intervention
- Lack of context for judging work

What are some of the benefits of exams (assumes quantitative)?

- Helps identify level of knowledge or skill (achievement and accountability)
- Results are easily quantified
- Individual performances can be easily compared
- Helps determine if intervention has made a difference in knowledge or skill level

What are some of the limitations of exams?

- Validity issues – does it test the appropriate knowledge and skills
- Results can be influenced by attitudes
- Language or vocabulary can be an issue
- People can be very concerned with how test results will be utilized (especially adults)
- Adults sometimes resent taking tests, which typically have a negative connotation

Appendix D: More Information on Random Sampling

Sample size

How big should your sample be? Make your sample size as large as you can afford in terms of time and money. The larger the sample the more you can expect it to reflect accurately what you would obtain by testing everyone.

Rules of Thumb

Population Size	Sample Size
50 or less	50 or less
500 or less	approx. 200
1,000 or less	approx. 275
10,000+	approx. 350
U.S. population	2,000 to 4,000

From Fitz-Gibbon, C.T. & Morris, L.L. (1987). *How To Design A Program Evaluation*. Newbury Park, CA: Sage Publications.

Random Sampling Techniques

To select randomly you can:

- number everyone in the population and use a random numbers table to select individuals (e.g. Random.org)
- draw names from a hat, like a lottery.
- determine how many interviews/surveys you need from a list of program participants, then take your list and count off participants (every fifth or twenty-fifth, depending on the total number and the number you need) to select individuals.

When the population consists of a number of subgroups (or strata) that may differ in the characteristics that interest you (such as grade level, or number of years teaching, or place of residence), use **stratified sampling**, which involves identifying your strata of interest, then drawing a sample from each strata.

Whatever you do, be consistent. Consistent errors are easier to find and mitigate than inconsistent ones. For more about samples and sampling, visit Columbia University's Center for New Media Teaching and Learning e-Lesson Samples & Sampling .

References

Alkin, Marvin C. 2011. *Evaluation Essentials: From A to Z*. New York: Guilford Press.

Little, Priscilla, DuPree, Sharon, and Deich, Sharon. 2002. *Documenting Progress and Demonstrating Results: Evaluating Out-of-School Time Programs.* Harvard Family Research Project and The Finance Project. Cambridge, MA. http://76.12.61.196/publications/OSTlocalevaluation.pdf

NOAA Office of Education and Sustainable Development. *Designing Evaluation for Education Projects.* Washington, D.C. http://wateroutreach.uwex.edu/use/documents/NOAAEvalmanualFINAL.pdf

Patton, Michael Quinn. 2002. *Qualitative Evaluation and Research Metho*ds. 3rd ed. Thousand Oaks, CA: Sage Publications.

Patton, Michael Quinn. 2008. *Utilization-Focused Evaluation*. 4th ed. Thousand Oaks, CA: Sage Publications.

Rockwell, Kay and Bennett, Claude. 2004. *Targeting Outcomes of Programs: A Hierarchy for Targeting Outcomes and Evaluating Their Achievement*. Faculty Publications: Agricultural Leadership, Education & Communication Department. Paper 48. http://digitalcommons.unl.edu/aglecfacpub/48

Russ-Eft, Darlene and Preskill, Hallie. 2009. *Evaluation in Organizations*. New York: Basic Books.

Taylor-Powell, E., Jones, L., and Henert, E. 2002. "Section 7: Using Logic Models in Evaluation: Indicators and Measures," in *Enhancing Program Performance with Logic Models.* Retrieved from the University of Wisconsin-Extension web site: http://www1.uwex.edu/ces/lmcourse/

U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. 2010. *The Program Manager's Guide to Evaluation.* 2nd ed. Washington, D.C. http://www.acf.hhs.gov/sites/default/files/opre/program_managers_guide_to_eval2010.pdf

W.K. Kellogg Foundation. 2004. *Evaluation Handbook*. Battle Creek, MI. http://www.wkkf.org/knowledge-center/resources/2010/w-k-kellogg-foundation-evaluation-handbook.aspx

Wolf, R. L. 1978. *Whatever happened to the Giant Wombat: An Investigation of the Impact of the Ice Age Mammals and Emergence of Man Exhibit.* Washington, D.C.: National Museum of Natural History, Smithsonian Institute.