Designing Evaluation

FOR EDUCATION PROJECTS

Office of Education and Sustainable Development
NOAA

Designing Evaluation

FOR EDUCATION PROJECTS

Office of Education and Sustainable Development
NOAA

Acknowledgments

Evaluation is an integral and essential component of effective project design. This manual is intended to assist project managers as they develop and implement their evaluation plan. Much of the manual is based on the Project Design and Evaluation course developed by NOAA Coastal Services Center and the work of the following individuals:

Kate Barba, NOAA Estuarine Reserves Division

Sarah Ross, NOAA Office of Education and Sustainable Development

Chris Parsons, Word Craft

Brian Miller, Indiana-Illinois Sea Grant Extension

Mike Spranger, Florida Sea Grant Extension

Ginger Hinchcliff, NOAA Coastal Services Center

Lynne Hinkey, NOAA Coastal Services Center

Corey Riley, NOAA Estuarine Reserves Division

In addition, discussions with other colleagues have shaped this document. Their assistance is appreciated.

Bora Simmons
Environmental Education
Department of Teaching and Learning
Northern Illinois University
DeKalb, IL 60115

Contents

1 Introduction 1

2 What is Project Evaluation? 1

3 Why is Evaluation Important to Project Design and Implementation? 2

4 Planning Evaluation 4

5 Types of Evaluation 9

6 Data Collection Methods 15

7 Mixed Methods 31

8 Ethics 32

Appendix A – Writing SMART Objectives 33

Appendix B – Logic Model 34

Appendix C – Minimizing Possible Errors in Random Sampling 35

Appendix D – Determining Sample Size – Rules of Thumb 35

Appendix E – Selecting the Right Data Collection Instrument 36

Appendix F – Using Rubrics for Scoring in Evaluation 38

Glossary 39

Selected References 40

1

Introduction

A considerable amount of time, effort, and other resources go into the development and implementation of education projects. Quite obviously, the goal is to create effective projects that can serve as models of excellence. Whether the project is an hour-long endangered species walk, a family day festival, marine resources monitoring, or a community forum, the aim of providing quality educational experiences remains the same.

This guide has been developed to help managers take project development to a new level by truly integrating evaluation into the process. The guide walks through the basics of evaluation, providing discussions on everything from types of evaluations and ways of collecting information to using an outside evaluator and the ethical considerations of gathering data from program participants. This information is intended to answer questions about project evaluation and provide guidance in using evaluation as a project improvement tool.

2

What is Project Evaluation?

In the course of implementing a project, various types of information are gathered. Project managers often want to know how many individuals participated in an event, whether participants were satisfied with the logistics, and whether staff members and volunteers feel confident in their ability to deliver a particular educational experience. Answers to these questions provide useful information. They help managers monitor specific aspects of the project. However, in practice, this type of information gathering tends to be more sporadic and patchy than methodical and comprehensive.

Evaluation is the systematic collection of information about activities, characteristics, and outcomes of projects to make judgments about the project, improve effectiveness, and/or inform decisions about future programming (adapted from Patton, 1987). Evaluation is not merely the accumulation and summary of data and information about a project. Importantly, evaluation provides managers with well-documented and considered evidence to support the decision-making process.

Project evaluation serves two general purposes. Evaluation helps to determine the project's merit (does it work?) and its worth (do we need it?). Evaluation helps decision-makers determine if a project should be continued and, if so, suggests ways to improve it. Additionally, evaluation documents project (and program) accomplishments. If the project has been designed properly with well-articulated objectives that specify what must be accomplished, to what degree, and within what time period, the evaluation can determine whether or not the objectives are being met. The evaluation can gather information as to why a project is or is not meeting its objectives.

3

Why is Evaluation Important to Project Design and Implementation?

The reasons for conducting a project evaluation vary as dramatically as do the projects and their contexts. Perhaps the most common reason revolves around the desire to understand, in a systematic way, what is and is not working in a project. All too often, how projects work (or don't work) is understood primarily through a combination of instincts, anecdotes, and numbers on a balance sheet. The passionate instructor who is wrapped up in the moment of teaching is probably not in a position to make defensible claims about the long-term impact of a lesson on a learner's education. Likewise, one cannot generalize from a series of anecdotes, however informative and even satisfying. Evaluation provides perspective. It provides evidence. It provides the types of information necessary for sound decision-making. The following outlines additional benefits of conducting project evaluations.

- **Participants** – Participants are core to the success of the project. In the long run, project sustainability will depend on the degree to which participants benefit directly, short-term and long-term, from the experiences or services. The evaluation will provide evidence of the ways in which participant learning is impacted.
- **Project Improvement** – Project strengths and weaknesses can be identified through an evaluation. Of equal importance, an evaluation can map out the relationships among project components – that is, how the various parts of a project work together. This information can be used to re-design the project and increase both efficiency and effectiveness.
- **Public Relations** – Data generated by an evaluation can be used to promote the products and services of the project within and outside of the agency. Instead of vague claims and uncertain assertions, statements based on evaluation results will be viewed as more substantial and justifiable. Importantly, if designed with sufficient care, an evaluation should be able to shed some light on why and how the project works.

- **Funding** – More and more program managers require the design and implementation of a comprehensive, outcomes-based evaluation. They want to know what types of impacts the project has had. Even if evaluation is not required, an evaluation can provide evidence of project effectiveness; such evidence may be important when limited resources are being distributed internally. Evaluation results are often helpful in determining if a project should be continued, scaled backed, discontinued, or enhanced.
- **Improved Delivery** – Projects evolve over time. What was once a coherent, discrete set of activities may have grown into a jumbled set of loosely related events. An evaluation can help clarify the purposes of the project, allowing decision-makers to examine project components against well-thought-out criteria. Valid comparisons of projects and activities can be made and duplication of efforts can be limited. It is quite possible that the evaluation will uncover a gem hidden in the jumble.
- **Capacity Building** – Engaging staff members, volunteers, and other stakeholders in the design and implementation of an evaluation will provide opportunities for skill building and learning. As the project or program is examined, those involved will also develop insights into the workings of the project and perhaps even the workings of the organization. These insights can be used to inform a strategic examination of projects and programs by identifying priorities, overlap, gaps, and exemplars.
- **Clarifying Project Theory** – When the project was designed initially, it was based either explicitly or implicitly on a project theory that explained how things work or how people learn or even how organizations change. An evaluation asks those involved to revisit that project theory. Based on experiences with the project and information taken from research literature, the evaluation provides an opportunity to revise the project theory. By making the project theory explicit, the underpinnings of the project and what makes it work will be better understood and thus, better implemented. Staff members and volunteers who understand why a particular set of teaching methods was selected or why the project activities were sequenced the way they were will be more likely to actually follow the plan. They will also feel more ownership in the project if they understand the theory behind the project more fully.
- **Taking Stock** – Engaging in evaluation provides a precious opportunity to reflect on the project. It is an opportunity to document where the project has been and where it is going, and consider whether the project is doing what its designers hoped it would do. Taking stock is more than accumulating information about the project, it is learning through the project.

4

Planning an Evaluation

As mentioned earlier, evaluation is the *systematic* collection of information about activities, characteristics, and outcomes of projects to make judgments about the project, improve effectiveness, and/or inform decisions about future programming. Consequently, great care must be made in the planning of any evaluation effort. The following outlines 12 steps of project evaluation. The outline is intended to break down a complex process into manageable steps. Please recognize, however, that in providing an overview of the process, nuances and detail are necessarily omitted.

Stage I: Planning

Step 1. Reiterate the issue, audience, and project objectives

Before a project evaluation can be designed, it is essential to understand fully the project – its components, the relationships among the components, the audience(s), and the intended outcomes (short-term, medium-term, and long-term). The project logic model should be examined and used as a roadmap for evaluation planning. With the logic model and the associated performance objectives in hand (see Appendix A – Writing SMART Objectives), evaluation planners will be able to articulate how the project is supposed to work. From this foundation, the rest of the evaluation design will follow.

Step 2. Establish planning team (including stakeholders, audience, and evaluators)

The project, in all likelihood, involves a variety of players. Project managers, resource managers, staff members, volunteers, participants, and community members all have a stake in the overall success of the project. Each plays a different role and sees the project through a different lens. These perspectives should be tapped when planning an evaluation. To ensure that ideas and perspectives are represented, members of stakeholder groups should be invited to participate in an evaluation planning team. The team, depending on the particulars of the evaluation, may play a purely advisory role or may take a more hands-on role in the actual data collection. The exact expectations of planning team members need to be decided and articulated early on in the process.

Step 3. Establish the goals and objectives of the evaluation

The project logic model (see Appendix B), created during the early phases of project planning, illustrates the relationships among the various project components (initial situation, identified priorities, inputs, outputs, and short-term, medium-term, and long-term outcomes). The logic model should help guide project evaluation, by serving as a tool for clarifying program elements, on-going self-evaluation, and identifying evaluation questions and indicators. At some point, however, the evaluation team will need to determine the scope of the evaluation – that is, to define the purpose of

the evaluation, what is going to be evaluated, and who will use the evaluation. Given time and resource constraints, it may not be realistic to expect that all aspects of the project will be evaluated. The project team will need to set specific goals and objectives that can be used to focus evaluation planning and design. A careful consideration of the project theory should be included in this discussion. To get the most out of an evaluation process, the project theory should be made explicit and put to a test. In doing so, the evaluation should be able to shed light on deeper questions related to how and why the project works the way it does.

Step 4. Clarify the time frame in which the activities and impacts (outcomes) are expected to occur

The relationships between overall project design and implementation, and project evaluation design and implementation cannot be over-estimated. Project evaluation is integral to project design. The evaluation time line must be integrated into the project implementation time line and visa versa. Otherwise, important opportunities, such as the ability to collect baseline data or assess critical impacts, can be easily missed and lost forever. Planning and implementing an evaluation can take several months. Sufficient care must be given to the development of the evaluation time line to ensure effectiveness.

Step 5. Perform a literature search (to establish benchmarks or standards; if appropriate)

Evaluation, like overall project planning, rarely takes place in a vacuum. As mentioned previously, projects are developed based on explicit or implicit theories of how the world works. In designing an evaluation it is helpful to consider the related literature and use this literature as a touchstone. Likewise, research into evaluation processes, practices, and standards is useful. This is particularly true if the evaluation team does not include an outside evaluation expert or someone with significant evaluation expertise. Finally, existing sources of information (previous evaluations, census data, reports, budgets, etc.) should be tapped.

12 Steps of Planning a Project Evaluation

Stage I: Planning

1. Reiterate the issue, audience, and project objectives
2. Establish planning team (including stakeholders, audience, and evaluators)
3. Establish the goals and objectives of the evaluation
4. Clarify the time frame in which the activities and impacts (outcomes) are expected to occur
5. Perform a literature search (to establish benchmarks or standards, if appropriate)
6. Select data collection methods and develop questions based on the evaluation goals and objectives

Stage II: Data Collection

7. Determine the audience sample
8. Design and pilot data collection instrument
9. Gather and record data

Stage III: Data Analysis and Reporting

10. Perform data analysis
11. Manage data
12. Synthesize information and create report

Step 6. Select data collection methods and develop questions based on the evaluation goals and objectives

By this point in the process, the evaluation team has determined why an evaluation is being conducted, who will conduct it, what will be evaluated (including defining project theory, expected outcomes, and indicators), who will be evaluated, and when the evaluation will take place. Each of these decisions begins to define the type of evaluation (e.g., formative, summative) to be conducted (see below) as well as the data collection tools (e.g., survey, interview, case study, focus group) that are most appropriate (see discussion of tools).

Stage II: Data Collection

Step 7. Determine the audience sample

Except in rare cases when a project is very small and affects only a few participants, evaluations will be limited to a subset of the total anticipated audience. The preferred method for selecting the subset is random sampling – using procedures that will reduce sample bias and response bias by selecting a sample that accurately reflects the population (see Appendix C). A sample represents the population if every person in the population has an equal chance of being selected. In general, to reduce sampling errors, make the sample as large as feasible in terms of time and money. The larger the sample, the more generalizable to the population – that is, it more accurately reflects what would be obtained by evaluating everyone in the population.

In determining audience sample, some basic questions will need to be answered: How many audiences do you wish to sample? (For example, a project might identify teachers, parents, and students as its audiences. Each audience would be sampled independently.) How many individuals do you hope to assess? (See the Determining Sample Size – Rules of Thumb in Appendix D as a guide.) How can you best reach the audience(s) to collect data? (Visitors to a marine sanctuary might only be accessed during a short window of opportunity; members of an advisory group might be more readily available.)

Step 8. Design and pilot data collection instrument

Just as the initial design of the project required careful design and pilot testing of instructional materials to see how they worked, the data collection methods or tools (e.g., interview, focus group, survey, observation) need to be crafted and pilot tested. The evaluator will need to answer a series of questions that will help establish the nature of the data collection instrument(s), such as: How important is statistical precision? How important is

Sample Evaluation Timeline for the Development and Analysis of a Survey

Month 1	Form evaluation team; set expectations; conduct background research (literature review, gather existing data, review logic model)
Month 2	Determine survey method; determine sampling method; develop draft survey
Month 3	Pilot test survey and revise
Months 4–5	Administer survey (mail, phone, in person)
Month 6	Code and enter data; analyze data
Month 7	Write evaluation report and send results to stakeholders

in-depth information about one or more aspect of the project? How will data collection be standardized? Is contact information on target audience(s) available? What method would the audience be most receptive to? How many questions can be asked?

Step 9. Gather and record data

Again, just as the design of a project requires the consideration of various logistics (e.g., staff schedules, availability of facilities), the data collection process must be thoroughly scoped. The evaluation team will need to determine how data will be collected (Who will conduct interviews or administer questionnaires? How will data be recorded?). Depending on the expertise of those who will be responsible for data collection, training of interviewers, focus group facilitators, or observers may be appropriate. Additionally, evaluators will need to design a system that assures anonymity and that all aspects of ethical standards are practiced.

Stage III: Data Analysis and Reporting

Step 10. Perform data analysis

Analyzing quantitative and qualitative data is often the topic of advanced research and evaluation. It is always a good idea to include a planning team member with survey and statistical analysis and/or qualitative data analysis expertise on the evaluation team. When that isn't possible, there are a few basics that can help in making sense of the data.

a. Have a plan in place for how to analyze, synthesize, store, and manage data *before* starting the data collection.

b. Develop a plan to guarantee an unbiased analysis and reporting of the data.

c. Always start analyzing the collected data with a review of the evaluation goals and objectives. (Why was the evaluation undertaken? What question did you want to answer?)

d. Make copies of all data, and store a master copy of the original data in a safe place.

e. For qualitative data, anticipate responses (the pilot test will help with this), and have a plan for categorizing and coding the data.

Step 11. Manage data

After the data are collected and even after the data have been analyzed, a plan must be put in place to continue the effective and ethical management of the data. In most cases, the data remain viable for some period of time. After reading the evaluation report, decision-makers, other stakeholders, and other evaluators may generate questions that can be answered by revisiting the data. Consequently, it is important to develop a plan for continued access to uncorrupted data. If data are to be retained for some period of time, the project must also make certain that the confidentiality and anonymity of respondents is maintained. Finally, intellectual property rights need to be defined (for example, who "owns" the data?).

Step 12. Synthesize information and create report

After the data have been collected and analyzed, an evaluation report must be written. Obviously, knowing when the evaluation report and its recommendations will be needed impacts the overall time line as well as decisions related to evaluation design and data collection strategies. Although the nature of the evaluation report (e.g., length, style of writing) is determined to some extent by its intended audience, there are

Guidelines for Conducting a Successful Evaluation

1. Invest heavily in planning.
2. Integrate the evaluation into ongoing activities of the program.
3. Participate in the evaluation and show program staff that you think it is important.
4. Involve as many of the program staff as much as possible and as early as possible.
5. Be realistic about the burden on you and your staff.
6. Be aware of the ethical and cultural issues in an evaluation.

From: U.S. Department of Health and Human Services. *The Program Manager's Guide to Evaluation.*

standard components to any evaluation report. For example, the report must include a description of the evaluation methods used – how was the sample drawn? How were data collected and analyzed? In addition, the report must include a discussion of the problems encountered and errors made in the design and implementation of data collection tools. Readers of the report must feel confident that the results of the evaluation are credible and the recommendations sound.

Who Should Conduct the Evaluation?

Early on in the evaluation planning process, the decision will need to be made whether or not to hire an outside evaluator. In some cases, the decision may have been made for the team. Many program managers require that an outside evaluator be contracted. An outside evaluator is seen as an objective third-party who may be able to bring a fresh perspective to the project. In addition, a professional evaluator has expertise that may not exist in-house. If evaluation expertise does exist and an outside evaluator is not required, it is possible to run major components of the evaluation in-house. However, managers should proceed with caution. Those with strong ties to the project (e.g., project manager, staff members, volunteers, advisory committee members) may find it difficult to shed their biases, particularly if evaluation results are to be used in decision-making. Project staff members, volunteers, and other stakeholders should be involved in the evaluation. They should play a role in determining both the focus and objectives of the evaluation. However, at a minimum, an outside evaluator should be responsible for data analysis and interpretation.

Evaluation Cost

Depending on the size and complexity of the project, it is typically recommended that 5-15% of project costs be set aside for evaluation. Costs may be lowered if some or even most of the evaluation is conducted in-house. (Remember, however, that even though a check may not be written to an outside evaluator, staff time still costs.) It should also be remembered that a great deal of data may already exist. For example, gleaning data collected for yearly reports may prove practical. Participant exit interviews may contain valuable information. Obviously, building an evaluation around existing data has its pitfalls, but a critical examination of quality sources of existing information that helps answer the evaluation questions should not be overlooked.

5

Types of Evaluation

As with many ventures, there are several different types of evaluation. Depending on the stage of project planning, managers may conduct a needs assessment (sometimes referred to as "front-end evaluation"), formative evaluation, or summative evaluation.

Front-End Evaluation (Needs Assessment)

As might be expected, a front-end evaluation or needs assessment takes place *prior* to undertaking the project. Front-end evaluation:

- Gathers information/data about the gap between the current and desired level of audience skills, knowledge, attitudes, and behaviors.
- Takes place *before* the project is designed.
- Helps confirm or negate assumptions of audience characteristics and appropriate content, define goals and objectives, and identify stakeholders and potential collaborators.

Questions that might be addressed by front-end evaluation include:

- What are the nature and scope of the problem? Where is the problem located, whom does it affect, and how does it affect them?
- What is it about the problem or its effects that justifies new, expanded, or modified projects or programs?
- What feasible actions are likely to significantly ameliorate the problem?
- What is the appropriate target audience(s)?

Formative Evaluation

Formative evaluation is conducted throughout the project, during project design, development, and implementation. Most often, formative evaluation is used to test out methods and materials. At its best, formative evaluation can be an essential decision-making tool that will transform the project. Formative evaluation:

- Gathers information/data about an audience's reaction to and learning from a project's pilot or prototype products/materials. Changes may be made as a result of formative evaluation.
- Gathers information/data about problems with project delivery, and assesses progress towards outcomes of a project during implementation.

- Helps provide information that can be used in making decisions about modification, continuation, or expansion of the project. (Results of a formative evaluation may be used to decide how to move forward with an existing project.)

Questions that might be addressed by formative evaluation include:

- Is a particular project reaching its target audience?
- Is the project being implemented well? Are the intended activities, products, or services being provided?
- Is the project effective in attaining the desired objectives or benefits?

Summative Evaluation

Again, as might be expected, summative evaluation is typically conducted at the *end* of the project. A formal report is developed that can be submitted to decision-makers and other stakeholders. Summative evaluation:

- Gathers information/data about the audience's skills, knowledge, attitudes, and behaviors at some point in time *after* project implementation begins. Results of a summative evaluation are used to inform stakeholders about the value of a project.
- Informs decision-makers about the value or worth of the project.
- Helps provide the information necessary to make decisions about the continuation, revision, or expansion of the project. (Results from a summative evaluation are used to determine if the project was "successful.")

Questions that might be addressed by summative evaluation include:

- Did the project reach its target audience?
- Was the project implemented well? Were the intended activities, products, or services provided?
- Was the project effective in attaining the desired objectives or benefits? How did the project impact the intended audience(s)?
- How much did the project cost?
- Is the project cost reasonable in relation to its effectiveness and benefits?

Levels of Evaluation

There are a number of levels on which to evaluate projects. Each of these levels provides slightly different information about the impacts of a project, from the simplest and most immediate (reaction) to the long-term, most complex (return on investment).

It is important to note that although these methods are presented in a hierarchy of increasing complexity, the hierarchy does not indicate relative value. All of these levels are useful. The choice of level(s) should be determined by the type of information needed to evaluate the project accurately.

Five Levels of Evaluation

Level 1. Reaction – What is the participants' response to the project or activity?

Level 2. Learning – What did the participants learn?

Level 3. Behavior or Application – Did the participants' learning affect their behavior?

Level 4. Results – Did participants' behavior change move the original situation towards the objective (desired outcome)?

Level 5. Return on Investment (ROI) or Cost-Benefit – Is the cost of implementing this project reflected in the level of benefits received from the results?

Source: Kirkpatrick (1994). The 5th level, Return on Investment, is often included among the levels of evaluation, although the original Kirkpatrick model does not include it.

Level 1. Reaction

Reaction evaluation measures the audiences' immediate positive or negative response to the project or learning experience. This is the most common level of evaluation.

Often referred to as "smile sheets," these evaluations ask participants to rate their perceptions about the quality and impact of the specific project or activity. Smile sheets can range from a handful of questions regarding the project delivery, facility, and/or usefulness, to forms that ask participants to rate all aspects of the activity.

Reaction evaluations are an important tool to measure participants' satisfaction. They are relatively easy to administer, tabulate, and summarize in a results report.

Example Questions for Level 1 – Reaction Evaluation:

Attending the workshop was:

Poor use of my time 1…2…3…4…5 Good use of my time

Length of workshop in relationship to the materials presented was:

Too long 1…2…3 (just right)…4…5 Too short

Workshop facilities were:

Inadequate 1…2…2…3 (adequate)…4…5 Great

What were the strengths and weaknesses of the workshop?

Level 2. Learning

Learning evaluation measures whether participating in the project increases the audience's knowledge and awareness of the issues addressed.

A number of different tools can be designed to measure what project participants have learned. Before and after tests, simulations or demonstrations, or other in-class (observable or measurable) methods allow instructors or project designers to determine if the knowledge or skills identified in the objectives were learned. It is important to remember that regardless of the method used to determine this, the "test" must relate directly to the course objectives.

Learning evaluations are more difficult to design and administer than reaction evaluations. One reason is that they must be customized for every instructional activity or project and must reflect the conditions of the specific job or real-world application of the learning. It is also important to remember that learning evaluations measure the level of knowledge or skills of participants at the time the test is administered. These evaluations do not measure long-term knowledge or skill retention, nor are they an indication of how these will be applied to a real-world situation.

Example Questions for Level 2 – Learning Evaluation:
What was the most important thing you learned by participating in the workshop?

List three benefits of an estuary:
1.
2.
3.

True or False: An estuary is the same thing as a bay.

Level 3. Behavior or Application

Application evaluation measures if the participant has been able to use the new knowledge and skills learned.

Application evaluation is significantly more complex than the first two levels in that it requires contacting participants after they have had time to apply the new knowledge and skills. As with other evaluation levels, many different tools can be used to collect the data. Each tool has different strengths and limitations. Tools include surveys, interviews, focus groups, observations, and written document review. Regardless of the tool, the questions should present specific skill and knowledge areas, and ask participants if and how they have applied them since learning them. Questions should focus on relevance of the project, whether participants have gone back and used materials provided by the learning experience, how new knowledge has been applied, and use of new skills.

Measuring the application of new knowledge and skills learned is becoming more accepted as a level that should be evaluated. It is important to know not only that participants understood the material during the learning experience, but that they were then able to go back to their homes, communities or jobs and apply it. This level of evaluation provides evidence of whether transfer of learning has occurred. It is much more powerful to justify a project by demonstrating that participants used the information

rather than reporting the number of participants that "liked" the project. Many decision-makers are now demanding increased accountability of resources spent educating the target audience.

Example Questions for Level 3 - Application Evaluation

Have you applied the skills you learned at the workshop to your current projects?

Not at all 1…2…3…4…5 Extensively

Have you referred to the materials in the past 6 months?

Not at all 1…2…3…4…5 Extensively

Have you implemented the action plan developed at the workshop?

Not at all 1…2…3…4…5 Extensively

Please describe one way you have used the materials in the past 6 months:

Describe one component of your action plan that you have implemented fully:

Have there been barriers in applying the information learned during the workshop?

No___ Yes ___ Please explain your answer:

Level 4. Results

Results evaluation measures the degree to which any behavioral changes have impacted on the environment or audience's lives.

There is constant pressure within agencies to demonstrate the efficiency and effectiveness of their programs. In order to actually conclude that a project has had its desired effect, participants have to "successfully" apply the new skills or knowledge. That is, the application of new skills and knowledge leads to the desired result or impact on an audience or the environment. This level of long-term feedback is becoming increasingly important particularly when priorities are being set or when decisions to continue or discontinue the project are being made.

Results evaluations are typically feasible only for large scale projects designed to produce specific results for a specific audience. For example, if you were trying to measure the results of teaching participants how to be facilitators, you would need to go to the people who have experienced facilitation conducted by the project participants. This requires that measurements be at least one step removed from the initial participants in your project. Because it can be quite difficult to isolate the effect of your project, this level of evaluation can be complex.

Example questions for Level 4 – Results Evaluation:

—After training on estuary restoration:

How many acres of estuaries have been successfully restored?

—After staff members participate in customer service training, customers would be asked:

> How satisfied are you with the level of service provided by XXX?
> Not satisfied 1…2…3…4…5 extremely

Level 5. Return on Investment (ROI), or Cost-Benefit Analysis

Cost-Analysis Approaches in Evaluation

Cost-benefit analysis (CB)
Cost-benefit is the evaluation of a project/product according to its cost and benefits when each is measured in monetary terms. A project/product must show benefits in excess of costs. In selecting from among several alternative projects, you would choose the one that had the highest benefit-cost ratio (or, conversely, the lowest ratio of costs to benefits.) For example, let's say you produce a CD on nonpoint source pollution. You track the total costs of production, marketing, and distribution, and you estimate the benefits that result from distributing the CD. From those data you develop a benefit-cost ratio and assess the value of that product based on the ratio. (Note: The major disadvantage to cost-benefit is that it is very difficult to assign monetary value to benefits.)

An alternate, more commonly-used approach is:

Cost-effectiveness analysis (CE)
Cost-effectiveness is the evaluation of alternative projects/products according to both their costs and their effects in producing some outcome (Levin & McEwan, 2001). The most preferable alternative would be the one that shows the lowest cost for the desired changes in outcome. For example, let's say you produce two products on nonpoint source pollution: a Web site and a brochure with a CD. Outcomes are measured as the number of people receiving information. You track the total costs and outcomes of both products, then compare the two to determine which product provides the maximum effectiveness per level of cost (or, conversely, the least cost per level of effectiveness).

6

Data Collection Methods

Evaluators have an array of data collection methods available to them. Although selecting the most appropriate method requires thought and careful consideration, the selection process is also shaped by a number of decisions that have already been made. The type of evaluation being considered (i.e., front-end, formative or summative) will determine, to some extent, the most appropriate data collection tool(s). Likewise, the level of evaluation being conducted (e.g., learning, application, results), the audiences involved (e.g., children vs. adults, casual visitors vs. organized groups), and the amount of resources available (e.g., time, money) will all help determine which tools should be used. Each data collection strategy comes with strengths, weaknesses, and a set of considerations.

The following table summarizes the purpose, advantages, and challenges of using nine different data collection tools when conducting evaluations. Remember that since data gathering tools are developed for a specific purpose and project, they rarely represent a pure form. For example, a survey or interview may include test items. A case study often incorporates observation, document review, and in-depth interviews. For further comparison and use of tools, see Appendix E.

Uses, Benefits, and Limitations of Various Data Collection Methods for Evaluation

Methods	Overall Purpose	Advantages	Challenges
Interviews	To fully understand someone's impressions or experiences, or learn more about their answers to questionnaires.	• Provides full range and depth of information • Develops relationship with respondent • Allows for follow-up questions	• Can take much time • Can be hard to analyze and compare • Can be costly • Interviewers can bias responses • Generalization may be limited
Focus Groups	To explore a topic in depth through group discussion, e.g., about reactions to an experience or suggestion, understanding common complaints. Useful in evaluation and marketing.	• Quickly and reliably obtain common impressions • Can be efficient way to gather range and depth of information in short time • Can convey key information about projects	• Can be hard to analyze responses • Need good facilitator for safety and closure • Difficult to schedule 6-8 people together
Questionnaires and surveys	To quickly and/or easily obtain a lot of information from people in a non-threatening way.	• Can complete anonymously • Inexpensive to administer • Easy to compare and analyze • Can administer to many people • Can obtain lots of data • Many sample questionnaires already exist	• Might not get careful feedback • Wording can bias client's responses • Impersonal • In surveys, may need sampling and statistical expertise • Doesn't yield full story
Observation	To gather accurate information about how a project actually operates, particularly about processes.	• Can view operations of a project as they are actually occurring • Can adapt to events as they occur	• Can be difficult to interpret behaviors • Observations can be difficult to categorize • Can influence participant's behaviors • Can be expensive
Literature Review	To gather information on the audience and/or the issue. Identify what previous investigations have found about the knowledge, skills, behaviors, or attitudes of the intended audience with relation to the issue.	• Can provide much information in relatively little time • Has most likely been reviewed or seen by audience • Makes use of already gathered information • Helps to sort changes over time • Provides evidence about the problem • Minimum effort or interruption of audience	• Can be out-of-date (e.g., technology needs) • Data synthesis can be difficult • May not address specific questions of concern • Not a flexible means to get data; data restricted to what already exists • Statistical data may not address perceptions of the problem, or may not address causes of the problem • Reports may be incomplete

Uses, Benefits, and Limitations of Various Data Collection Methods for Evaluation (cont.)

Methods	Overall Purpose	Advantages	Challenges
Tests	To determine the audience's current state of knowledge or skill regarding the issue.	• Helps identify a problem or a deficiency in knowledge or skills • Results are easily quantified • Individual performances can be easily compared • Helps determine if the problem is a training issue	• Limited availability of validated test for specific situations • Results can be influenced by attitudes • Language or vocabulary can be an issue • People may be concerned about how results will be used • Adults may resent taking tests
Concept Maps	To gather information about someone's understanding of, and attitudes towards, a complex subject or topic.	• Can offer a more comprehensive and complex view of someone's thinking than a test does • Could be a better tool for visual learners or test-phobic people • Produces qualitative and quantitative data	• Takes training to complete properly • Takes training to administer • Can be challenging and time consuming to score • Can be difficult to analyze and interpret
Document or Product Review	To gather information on how the project operates without interrupting the project; comes from review of applications, finances, memos, minutes, etc.	• Yields historical information • Doesn't interrupt project or client's routine in project • Information already exists • Few biases about information	• Often takes much time • Information may be incomplete • Need to be quite clear about what looking for • Not a flexible means to get data; data restricted to what already exists
Case Studies or Peer Review	To fully understand or depict client's experiences in a project; to conduct comprehensive examination through cross comparison of cases.	• Fully depicts client's experience in project input, process, and results • Powerful means to portray project to outsiders	• Usually quite time consuming to collect, organize, and describe • Represents depth of information, rather than breadth • Information gathered cannot be generalized

Adapted from McNamara, C. *Basic guide to program evaluation.* (Retrieved from *http://www.mapnp.org/library/evaluatn/fnl_eval.htm* December 27, 2003.)

Interview

What is it?
Active interchanges between two people either face to face or via technology (e.g., telephone, email).

How many respondents/participants?
Typically 5–20. The number of individuals interviewed can be increased, but time and cost also increase.

Time issues?
Interviews can last 15 minutes to 2 hours, 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
- Unskilled interviewers 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
- 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?
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 has 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

Questionnaire and 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?

Costs include printing, postage, return postage, follow-up postcards. (Costs increase with a large sample size.) If surveys are administered in person (either individually or in small groups), costs can diminish dramatically. Email surveys are least costly, but the number of questions that can be asked is limited.

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 (respondents complete and return)
- Large sample size; data can be generalized to population
- Range from inexpensive to expensive (depending on design and administration)
- 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 (for written instruments)

- Useful at all evaluation stages
- High level of return for interview-style and most group-style surveys
- Good for gathering information that requires sequencing (respondents can't read ahead if given as an oral survey)
- Survey administrators can clarify questions if conducted in person or over the phone
- 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

Observation

What is it?

Data collection based on watching a process or skill and systematically recording events. These observations may be made in person (at the time of the event) or using media (e.g., analysis of a video tape recording).

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 complexity of the subject.

When to use it?

Observation allows evaluators to document behavior. When evaluators want to know how people behave (e.g., demonstration of skills, recycling, fishing) or the results of specific behavior (e.g., diversity of plants or animals in a restoration area), observation should be used. Actual behavior (or the results of the behavior) is documented, not what people say they do or are planning on doing.

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 in the system
- Administrative costs can be kept to a minimum

What are some of the limitations?

- Requires process and content knowledge by observer
- Observer can disrupt or alter the system

- Observer can be considered a spy
- Data can be skewed by observer's biases (and skills)
- Data are not always quantifiable and may require judgments by the observer
- Typically, small sample size
- Usually time intensive
- Does not indicate how participants view their actions
- Replication difficult

Literature 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.

How many respondents/participants?
N/A

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
- Makes use of already gathered statistical data
- Easier to chart changes over time
- Provides excellent evidence of problem (front-end evaluation)
- Minimum effort or interruption of workers

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
- Reports may have been adjusted or “selectively edited”

Test

What is it?

An exam that assesses knowledge or skill level. Can be essay, fill in the blank, true/false, and/or multiple choice formats.

How many respondents/participants?

From 25 to thousands. Numbers are limited by the length of the test (think SAT or GRE) and its format. Scantron (or bubble tests) can be administered to hundreds or thousands (assuming access). Essay tests are limited by the ability to score them.

Time issues?

10 minutes–1 hour to administer (although most participants would resist a test that takes more than a few minutes to complete). Development of a valid and reliable test instrument can take months. Data analysis, if closed-ended questions are used, should take a short amount of time.

Cost issues?

Inexpensive to moderately expensive. Greatest cost issues revolve around the development of the test (ensuring validity and reliability). If instruments exist or simple measures of knowledge/skills are used, development costs will be limited.

When to use it?

Tests are used when evaluators want to assess the audience's level of knowledge or skills. Tests 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).

What are some of the benefits?

- 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?

- Limited availability of validated tests for specific situations
- 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

Concept Map

What is it?
A graphic representation of how an individual thinks about a topic.

How many respondents/participants?
Typically 5–20. The number of individuals asked to produce a concept map can be increased, but time and cost also increase.

Time issues?
5–45 minutes. Before data can be collected, respondents will need to learn how to draw a concept map. Data analysis is slow and time consuming (can take weeks, especially if concept maps are complex – showing many relationships among ideas – or if a large number of individuals produced concepts maps).

Cost issues?
Analysis of concept maps requires expertise and training.

When to use it?
Concept map is best used when the evaluator wants to understand the relationships among ideas or attitudes. Because it is structured differently from a test or typical survey, constructing a concept map may be less intimidating for some (especially for those who are uncomfortable with tests).

What are some of the benefits?

- Provides a visual representation of how individuals see the topic and the relationships among ideas
- Tends to reduce test anxiety (people often see it as a fun exercise rather than a test of understanding)
- Allows respondents to illustrate complex ideas and relationships (particularly useful for visual learners)

What are some of the limitations?

- Difficult to score (data analysis must be completed by someone with specific expertise and training)
- Time consuming to score (each idea and relationship must be analyzed)
- Participants must be taught how to construct a concept map
- Can be difficult to determine commonalities (patterns) among participants or compare participant understandings

Document or Product Review

What is it?
Systematic examination of documents (e.g., rosters, time sheets, portfolios, photographs, participant work) collected during a project.

How many respondents/participants?
N/A

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

Cost issues?
Data collection relatively inexpensive, assuming that documents can be retrieved easily. Data analysis may be time consuming and therefore expensive.

When to use it?
Documents can be used to quantify particular project components (e.g., number of participants, amount of time spent on task, costs). Document review is unobtrusive and does not impact on the delivery of the project. Documents can provide multiple measures of the same construct, allowing evaluators to examine the process from more than one direction.

What are some of the benefits?
- Relatively cost effective
- Assuming documents have been archived, data analysis can take place at any time
- Documents can be used to monitor the development of a project
- Project history can be documented
- Data collection is often unobtrusive

What are some of the limitations?
- Collection and analysis of documents can be time consuming (especially if documents have not been archived in an orderly manner)
- Available information may not match evaluation questions and needs

Case Study

What is it?
Holistic, detailed narrative describing a project or the experiences and perceptions of a particular participant.

How many respondents/participants?
One (in some cases 2–3).

Time issues?
Time intensive. Data collection, typically using ethnographic data collection methods, takes place consistently and continuously over an extended period of time. For the most part, data are qualitative and require a great deal of time to collect, organize, and analyze. May require the evaluator to be a participant observer.

Cost issues?
Data collection and analysis require a significant amount of time and expertise. Evaluator time could be costly.

When to use it?
Case studies are particularly useful when decision-makers or other stakeholders would like a detailed understanding of the project or the experience of a project participant, and are willing to forego generalizability. Case studies provide a rich, in-depth description of a participant or a small group of participants, or a project and how it works. Case studies provide vivid imagery, pulling a variety of data gathered from interviews, documents, and observation into a cohesive whole. Patterns of experience can be highlighted. Case studies can help decision-makers understand project successes and failures within a well-described context.

What are some of the benefits?
- Provides a holistic understanding of the project
- Provides vivid imagery (narrative, paints a picture of the project)
- Produces detailed understanding of patterns and experiences

What are some of the limitations?
- Time intensive
- Detailed understanding of a particular participant; results cannot be generalized
- Strategies for data collection and analysis are complex

Matching Data Collection Tools to What is Being Assessed

It is completely reasonable to expect that an evaluation would want to address the degree to which participants increased their level of understanding, developed a particular set of skills, or further considered their attitudes on a topic. Some data collection tools are particularly adept at assessing knowledge gain or behavior change. Others are appropriate for documenting skill or attitude development. The issue, as with much of project evaluation, is selecting the appropriate tool. The following chart provides some ideas of the appropriateness of using specific kinds of data collection methods to assess knowledge, skills, attitudes, and behavior.

Data collection method	Knowledge	Skills	Attitude	Behavior
Interview	X		X	(X)
Focus group	(X)		X	
Questionnaire and survey	X	X	X	(X)
Observation		X		X
Literature review*	X	X	X	X
Test	X	X		
Concept maps	X		(X)	
Document or product review	X	X	(X)	X
Case study	X	X	X	X

Notes:
(X) Indicates that this technique may be, but is not always, appropriate to evaluate the indicated type of learning.
* For comparison from past to initial condition.
Adapted from: American Society for Training and Development (1989). *Evaluation Tool Use*. Alexandria, VA: Author.

Validity and Reliability of Evaluation Tools

Validity and reliability are critical to the development and use of evaluation tools. The validity of an instrument is the extent to which it measures what it purports to measure. A test (or other instrument) may be valid for one purpose, but not another. The reliability of an instrument is the extent to which it measures a variable consistently time after time. An instrument can be valid and reliable, valid and not reliable, or reliable and not valid. Obviously, care must be taken to ensure that any data gathering methods are both valid and reliable.

7

Mixed Methods

Mixed methods, in evaluation, refers to the practice of using some combination of both quantitative and qualitative data gathering. Quantitative methods allow us to count events or number of participants, determine cost/participants, perform statistical analyses (mean, median, mode, standard deviation), and complete other calculations. Quantitative methods allow us to generalize the findings beyond the actual respondents to the relevant population.

Qualitative methods allow us to record explanations, perceptions, and descriptions of experiences – often in the participants' own words. Qualitative methods allow us to create narratives that provide an in-depth view and a more complete understanding of the context of the evaluation. Typically, a small number of individuals participate in a qualitative evaluation. Consequently, the results of this small number of participants cannot be generalized to the population.

Each method has its own strengths and weaknesses. Using quantitative methods or qualitative methods in isolation limits what can be learned from the evaluation, what can be reported, and what can be recommended, with any confidence, as a result of the evaluation. Used in combination, however, the individual strengths of quantitative and qualitative methods can be maximized and the weaknesses minimized. More importantly, a synergy can be generated when using mixed methods. Results from more than one method of data collection can be "triangulated," providing greater validity and enhanced understanding. A survey of participants may provide a great deal of information about what services are most desired (and least desired); an interview of a small number of the participants may then provide in-depth information concerning why those services are most desired (or least desired) and, importantly, what characteristics make a particular type of service most desired.

8

Ethics

Virtually all evaluations involve collecting some information, directly or indirectly, from individuals. In designing the evaluation, the project team must ensure that the individuals involved are treated with respect and sensitivity. Not only should the maintenance of confidentiality and/or anonymity be a high priority, but the time and effort expended by evaluation participants should be respected. It is not appropriate, for example, to collect data from individuals when there is no specific plan to use that data. Respondents' physical and psychological well-being must be assured throughout the data collection process. Whether they are providing demographic data on a survey, completing a test, or responding to an interview, respondents are disclosing personal information about themselves. Interviews, surveys, and other instruments must be designed in such a way that evaluation participants are not embarrassed or asked to do something that might put them in jeopardy. Respectful data collection is not enough, however. Evaluation ethics require that respondents understand that they are participating in an evaluation and give their permission (with minors, a parent or guardian must provide informed consent).

Appendix A

Writing SMART Objectives

Effective project development and implementation depends on the clear articulation of goals and objectives. While goals provide a vision and overall description of the project, objectives chart expected outputs and outcomes. In writing meaningful objectives, many project developers have found a set of criteria, summarized by the acronym SMART, to be helpful. A SMART objective is:

Specific – Describes an action, behavior, outcome, or achievement that is observable.

Measurable – Details quantifiable indicator(s) of progress towards meeting the goal (e.g., 70% of participants..., five or more...)

Audience – Names the audience (e.g., workshop participants, community members) and describes outcomes from the perspective of the audience (i.e., what the audience will be able to do).

Relevant – Is meaningful, realistic, and ambitious; the audience can (given the appropriate tools, knowledge, skills, authority, resources) accomplish the task or make the specified impact.

Time-bound – Delineates a specific time frame.

SMART objectives use specific action words that describe what the learner will be able to do as a result of participating in the project. Action words can help to categorize types of learning in a hierarchy ranging from simple to complex. For example, analysis is a different level of learning than is application. One of the best known tools for developing objectives is the 1956 publication *Taxonomy of educational objectives: The classification of educational goals. Handbook I. Cognitive Domain* by Benjamin Bloom and a team of educational psychologists. The *Taxonomy* categorizes six levels of learning that commonly occur in education settings: knowledge, comprehend, apply, analyze, synthesize, and evaluate. (See the following table for some of the verbs that could be associated with each of the six levels of learning.)

Examples of Action Words Used to Help Set Objectives for Different Levels of Learning

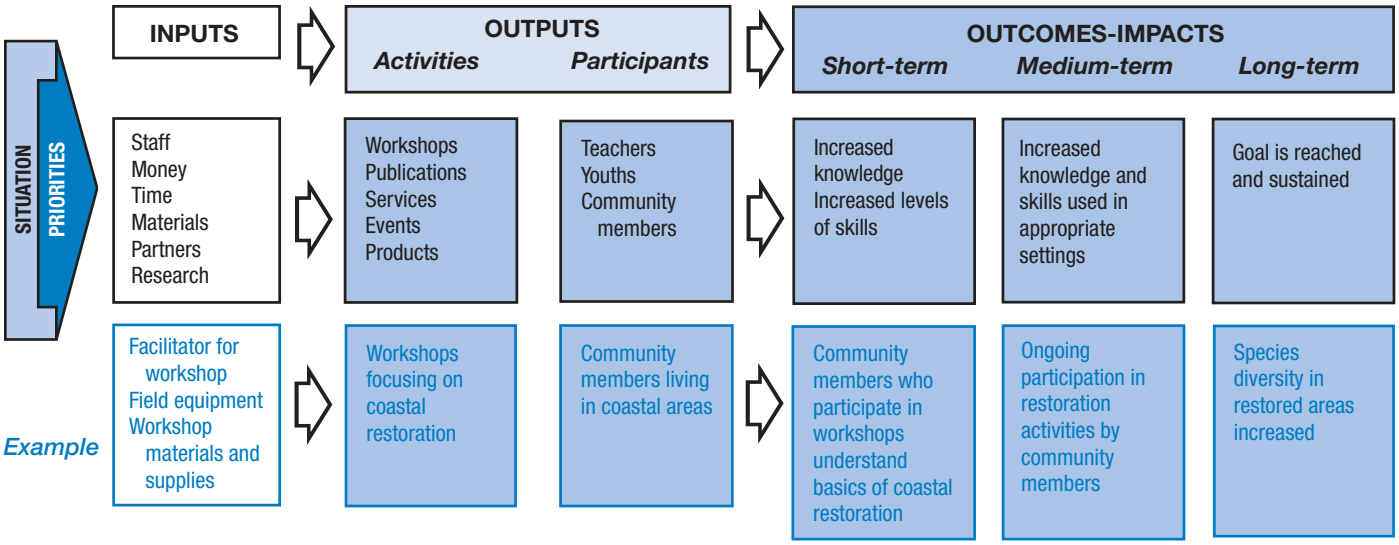
Knowledge	Comprehend	Apply	Analyze	Synthesize	Evaluate
define	discuss	demonstrate	distinguish	design	appraise
record	explain	employ	debate	construct	assess
list	differentiate	illustrate	calculate	create	judge
name	identify	translate	diagram	propose	predict

Appendix B

Logic Models as a Tool for Program Development and Evaluation

The logic model provides a visual representation of the program and its evaluation. The logic model illustrates the relationships among the various program components: initial situation (e.g., degraded coastal areas with declining numbers of species), identified priorities (e.g., restoring coastal areas, increasing species diversity); inputs (i.e., resources needed to accomplish a set of activities); outputs (i.e., activities designed to accomplish the program goal, as well as the audiences that participate in those activities); and short-term (immediate), medium-term (2-3 years), and long-term (4-10 years) outcomes-impacts. The logic model can help guide program planning, implementation, and evaluation. It can serve as a tool for clarifying program elements, identifying evaluation questions and indicators, and conducting ongoing self-evaluation.

Logic Model, Evaluation Questions and Indicators



Evaluation Questions: What do you want to know?

Were the inputs sufficient and timely? Did they meet the program goals?	Did all activities occur as intended? What was the quality of the intervention? Was the content appropriate?	Did targeted community members participate? Who did not participate? Who else was researched?	Did knowledge increase? Did understanding of coastal restoration techniques increase? What else happened?	Are community members continuing to participate in restoration activities? Are they participating in other activities?	To what extent has the biodiversity of the targeted coastal area been increased? In what other ways has ecosystem quality increased?

Indicators: How will you know it?

Number of staff; Funds invested; Delivery timetable	Number of workshops scheduled; Publications printed; Number of events	Number of workshops scheduled; Publications printed; Number of events	Number, percent with increased knowledge of coastal restoration; Additional outcomes: +, -	Number, percent using new knowledge and skills to monitor progress of restoration activities; Additional outcomes: +, -	Number of species recovered; Other positive environmental benefits; Additional outcomes: +, -

Adapted from Taylor-Powell, E., Jones, L. & Henert, E. (2002) "Enhancing Program Performance with Logic Models." Retrieved December 1, 2003, from the University of Wisconsin-Extension web site: www.uwex.edu/ces/pdande/evaluation/pdf/WaterQualityProgram.pdf.

Appendix C

Minimizing Possible Errors in Random Sampling

Type	Cause	Remedies
Sampling error	Using a sample, not the entire population to be studied.	Larger samples – these reduce but do not eliminate sampling error.
Sample bias	Some of those selected to participate did not do so or provided incomplete information.	Repeated attempts to reach non-respondents. Comparison of characteristics of non-respondents to describe any suspected differences that may exist.
Response bias	Responses do not reflect "true" opinions or behaviors because questions were misunderstood or respondents chose not to tell the truth.	Careful pre-testing of instruments to revise misunderstood, leading, or threatening questions. No remedy exists for deliberate equivocation in self-administered interviews, but it can be spotted by careful editing. In personal interviews, this bias can be reduced by a skilled interviewer.

Appendix D

Determining Sample Size – Rules of Thumb

Population	Sample
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

Source: Fitz-Gibbon & Morris. (1987).

Appendix E

Selecting the Right Data Collection Method

The tables below provide a convenient reference for the selection of appropriate evaluation data collection methods for different types of projects and activities. These tables provide general observations about the types of data gathering methods. The specific design and use of any particular method will be dependent on the particular context and the types of information needed.

A. Method versus Audience

Evaluation Methods* ***Audience***	**Interview**	**Focus Group**	**Survey**	**Observation**	**Test**	**Concept Maps**	**Case Study**
Adults/who know you or your organization	Good	Good	Good	Good	Good	Good	Good
Adults/who don't know you or your organization	Good	Good	Good		Fair to Poor	Good to Fair	Fair
Decision-makers/ policy makers/ community leaders	Good	Fair	Good to Fair	Good to Fair	Good to Fair	Fair	Good
Cultural groups (otherthan your own)	Fair to Poor	Good to Fair	Fair to Poor	Good to Fair	Poor	Good to Fair	Fair to Poor
Teachers	Good	Good	Good	Good	Fair	Fair	Good
Teens	Good	Fair	Fair	Fair	Fair	Good	Good
Eight to twelve year olds	Fair	Fair	Fair	Fair	Fair	Good	Fair
Three to seven year olds	Fair to poor	Fair to poor	N/A	Good	N/A	N/A	Fair to poor

Rating Scale: Good = offers more benefits than limitations; Fair = benefits and limitations are close to even; Poor = offers more limitations than benefits; N/A = tool in most cases is not appropriate for the audience. * Literature review and document review utilize existing sources of information that are not dependent on audience type.

Appendix E

B. Method versus Activity/Project

Evaluation Method*	Interview	Focus Group	Survey	Test	Observation	Concept Map
Project/Activity						
Talk/lecture (short, single event)	Poor	Poor	Fair	Good to Fair	Poor	Fair
Workshop (single event)	Fair	Fair	Good	Good	Poor	Good to Fair
Series (multiple meetings)	Good	Fair	Good	Good	Fair	Good
Training (skill building)	Good	Good	Fair	Fair	Good to Fair	Fair
Tour (adults)	Good	Fair	Fair	Fair	Fair	Fair
Tour (3-16 year olds)	Fair	Fair	Poor	Poor	Fair to Poor	Fair
Event/festival	Good	Fair	Good	N/A	Fair	N/A
Interpretive sign(s)	Good	Fair	Fair	Fair	Good to Fair	N/A
Exhibit	Good	Good	Good	Fair	Good to Fair	Good
Curriculum packet/materials	Good	Fair	Good	Fair	Fair	Good
Kits/activities	Good	Fair	Good	Fair	Fair	Good
Publications	Good	Fair	Fair	Fair	Poor	Poor
Media (e.g., video)	Good	Good	Good to Fair	Fair	N/A	Fair
Interactive media (e.g., CD)	Good	Good	Good to Fair	Fair	Good	Good
Website	Good	Fair	Good to Fair	Fair	Good	Good

Rating score: Good = offers more benefits than limitations; Fair = benefits and limitations are close to even; Poor = offers more limitations than benefits; N/A = tool in most cases is not appropriate for the project or activity. * Literature review and document review utilize existing sources of information that are not dependent on audience type. Case study involves creating a detailed description of a project or a participant's experiences with a project. Data for a case study can be collected through a combination of methods. The appropriateness depends, consequently, on the combination of methods used.

Appendix F

Using Rubrics for Scoring in Evaluation

Designing meaningful ways of scoring performance is essential to the development and implementation of effective education projects. This seems to be particularly true when we wish to gauge whether expected learner outcomes have been achieved. Many educators have found rubrics to be a useful tool when evaluating learner outcomes. Rubrics can be thought of as a logical extension of project and instructional objectives. They provide a template for scoring performance on evaluation activities such as portfolios, open-ended questions, and group or independent research. With a well-written rubric, it is reasonable to expect that all performances will be measured with the same yardstick. Additionally, when rubrics are used in educational contexts, learners know what is to be expected of them.

Sample Rubric – Taking Water Samples

Score	**Safety:** Degree to which learner follows correct safety procedures	**Procedures**: Degree to which learner follows proper mechanics in water quality analysis	**Results**: Degree to which learner obtains proper sample values	**Interpretation**: Degree to which learner develops likely hypotheses
4 Fully meets standards	Handles chemicals and glassware safely	Obtains uncontaminated samples and follows correct steps for pH analysis	Both samples within .3 points of the correct pH	Can list three plausible reasons why the pH of the two samples differs and can defend reasoning behind hypotheses
3 Partially meets standards	No serious safety issues during analysis, but procedures deviate from ideal	Has some problems following instructions, but procedure adequate for approximate correct test results	One sample within .3 points of the correct pH	Can list two plausible reasons why the pH of the two samples differs and can defend reasoning behind hypotheses
2 Major departure from some aspect of standards	Shows some concern or knowledge about safety issues, but is careless in handling materials	Major problems with procedures that will likely yield incorrect results	Neither sample within .3 points, but at least one sample within .5 points	Can list one plausible reason why the pH of the two samples differs and can defend reasoning behind hypothesis
1 Does not meet standards	Disregards safety concerns when handling materials	Does not follow necessary steps in analysis and cannot obtain useful results	Neither sample within .5 points	Cannot list even one plausible reason why the two samples differ

Glossary

Assessment involves gathering data (either formally or informally) to be used in forming judgments (adapted from Mehrens and Lehman, 1991).

Baseline Data are collected at the beginning of the project prior to any services or activities being conducted.

Evaluation is the systematic collection of information about activities, characteristics, and outcomes of projects to make judgments about the project, improve effectiveness, and/or inform decisions about future programming (adapted from Patton, 1987).

Gap Analysis is a method used to determine the gap between the current situation and the desired outcome.

Items are individual questions on an instrument.

Needs Assessment is a systematic investigation of an audience to identify aspects of individual knowledge, skill, interest, attitude, and/or ability relevant to a particular issue, organization goal, or objective

Population. The entire collection of individuals about whom one is trying to make accurate statements.

Qualitative Data are descriptive rather than enumerative. They are usually provided in the form of words, such as descriptions of events, transcripts of interviews, and written documents. Qualitative data can be transformed into quantitative data through coding procedures.

Quantitative Data are numeric data. Analysis of quantitative data involves looking at relationships between quantities.

Reliability is the extent to which a data gathering instrument measures a variable consistently time after time.

Response rates are the percentage of a selected sample from which data were collected (responses actually received). A further calculation can sometimes be made of the fractions of the population requested in the sample.

Sample. A subset of the population from which information is actually collected.

Survey Instruments are any consistent method or tool by which information is systematically gathered.

Validity. The validity of an instrument is the extent to which it measures what it purports to measure. A test may be valid for one purpose, but not another.

Selected References

Diamond, J. (1999). *Practical evaluation guide: Tools for museums and other informal educational settings.* Walnut Creek, CA: AltaMira Press.

Fetterman, D.M. (1998). *Ethnography: Step by step.* Thousand Oaks, CA: Sage Publications.

Fink, A. & Kosecoff, J. (1985). *How to conduct surveys: A step by step guide.* Newbury Park, CA: Sage Publications.

Fitz-Gibbon, D.T. & Morris, L.L. (1987). *How to design a program evaluation*. Newbury Park: Sage Publications.

Frechtilling, J. et al. (2002). *The 2002 user friendly handbook for project evaluation.* Washington, D.C.: National Science Foundation.

Herman, J., Morris, L.L. & Fitz-Gibbon, C.T. (1987). *Evaluator's handbook.* Newbury Park, CA: Sage Publications.

Kirkpatrick, D. (1994). *Evaluating training programs: The four levels.* San Francisco, CA: Berrett-Koehler.

Levin, H.M. and McEwan, P.J. (2001). *Cost-effectiveness analysis: Methods and applications*. Thousand Oaks, CA: Sage Publications.

Madison, A.M. (ed.) (1992) *Minority issues in program evaluation.* San Francisco, CA: Jossey-Bass Publishers.

McNamara, C. *Basic guide to program evaluation.* (Retrieved from http://www.mapnp.org/library/evaluatn/fnl_eval.htm December 27, 2003.)

Mehrens, W. & Lehman, I. (1991). *Measurement and evaluation in education and psychology* (4th ed.). Chicago, IL: Holt, Rinehart, and Winston, Inc.

Morris, L.L., Fitz-Gibbon, D.T. & Freeman, M.E. (1987). *How to measure performance and use tests.* Newbury Park, CA: Sage Publications.

Patton, M.Q. (1987). *How to use qualitative methods in evaluation.* Newbury Park, CA: Sage Publications.

_____. (2001). *Qualitative research and evaluation methods.* Beverly Hills, CA: Sage Publications.

Sanders, J. (1994) *The program evaluation standards – How to assess evaluations of educational programs.* 2nd Edition. Thousand Oaks, CA: Sage Publications.

U.S. Department of Health & Human Services. *Program manager's guide to evaluation.* (Retrieved from http:www.acf.hhs.gov/programs/core/pubs_reports/chapter_1_pmguide.html December 27, 2003).

Wholey, J., Hatry, H.P. & Newcomer, K.E. (Eds.) (1994) *Handbook of practical program evaluation*. San Francisco, CA: Jossey-Bass Publishers.