

FORREX SERIES 6

# Conducting Program and Project Evaluations: A Primer for Natural Resource Program Managers in British Columbia

# **Conducting Program and Project Evaluations: A Primer for Natural Resource Program Managers in British Columbia**

---

Murari Suvedi and Shawn Morford



## National Library of Canada Cataloguing in Publication Data

Suvedi, Murari

Conducting program and project evaluations [electronic resource] : a primer for natural resource program managers in British Columbia / Murari Suvedi and Shawn Morford.

(FORREX series ; 6)

Type of computer file: PDF.

System requirements: Adobe Acrobat Reader.

ISBN 1-894822-09-9

1. Forestry projects--Evaluation. 2. Forest management--British Columbia--Evaluation. I. Morford, Shawn, 1959- II. FORREX III. Title. IV. Series: FORREX series (Online) ; 6.

SD568.B7S88 2003

333.75'068'4

C2003-901705-2

© 2003 FORREX--Forest Research and Extension Partnership

Information may be reproduced without permission subject to the fair dealing provision and the exceptions set out in the Canada *Copyright Act*, R.S., c. C-20, s. 1. The source of the work must be fully acknowledged. Information may not be redistributed or stored for the purpose of serving through any other information retrieval system without the written permission of FORREX--Forest Extension Research Partnership. Links can be made freely. No guarantee or warranty, expressed or implied, is made about the value or stability of the information or links made herein. However, reproduction of this work, in whole or in part, for purposes of commercial use, resale, or redistribution requires written permission from FORREX--Forest Research and Extension Partnership. For this purpose, contact the Partnership at: Suite 702 235 1st Avenue, Kamloops, BC V2C 3J4

For more information, visit the FORREX Web site: <http://www.forrex.org>

This FORREX-Forest Research Extension Partnership publication is partially funded by Forestry Innovation Investment



The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by FORREX--Forest Research and Extension Partnership of any product or service to the exclusion of any others that may also be suitable.

## ABSTRACT

---

Increasingly, public and private agencies in the Canadian forest sector are being expected to demonstrate the effects of their programs, projects, and policies. This publication is designed to “demystify” program evaluation for natural resource professionals, and program and project managers. It covers the essential elements of program evaluation, to enable program leaders to talk meaningfully about evaluation and to conduct basic but meaningful evaluations. The program evaluation outcomes will allow program leaders to be more effective in improving programs, making decisions about the future of programs, and identifying cause-and-effect relationships between their activities and the outcomes of their work.

### *Citation—*

Suvedi, M. and S. Morford. 2003. *Conducting Program and Project Evaluations: A Primer for Natural Resource Program Managers in British Columbia*. FORREX—Forest Research Extension Partnership, Kamloops, B.C. FORREX Series 6.

## AUTHORS

---

*\*Correspondence to:* **Murari Suvedi**, Associate Professor  
Department of Agriculture and Natural Resource  
Education and Communications Systems  
College of Agriculture and Natural Resources  
409 Agriculture Hall  
Michigan State University  
East Lansing, Michigan 48824 USA  
E-mail: [suvedi@msu.edu](mailto:suvedi@msu.edu)

**Shawn Morford**, Socio-economics Extension Specialistt  
FORREX–Forest Research Extension Partnership  
506 Burnside Road  
Victoria, BC V8Z 1M5 Canada  
Tel: 250-363-0769  
Fax: 2250-363-3665  
E-mail: [shawn.morford@forrex.org](mailto:shawn.morford@forrex.org)

## CONTENTS

---

Abstract.....	iii
Introduction.....	1
Who Should Use this Publication? .....	1
What is Evaluation?.....	1
Determining the Purpose of an Evaluation .....	1
Types of Evaluations .....	2
Does Evaluation Involve Research? .....	2
Framing Your Thinking about Evaluation .....	2
Steps for Conducting Program Evaluation .....	3
When Should an Evaluation be Conducted? .....	4
Internal versus External Evaluators .....	5
Developing Clear, Specific, and Measurable Objectives .....	5
An Example of Program Evaluation .....	6
Reporting Evaluation Results .....	6
Where to Get More Information .....	7
Glossary .....	8
References .....	9
Appendix A: Program Evaluation: Methods of Collecting Data .....	11
Appendix B: Selecting a Sample in Quantitative Data Collection .....	16
Appendix C: Analyzing Evaluation Data .....	19
Appendix D: Establishing Validity and Reliability .....	20
Appendix E: Sample End-of-Workshop Evaluation .....	22

## **INTRODUCTION**

---

Increasingly, publicly funded agencies in Canada are expected to demonstrate the impacts of their programs, projects, and policies. The *British Columbia Budget Transparency and Accountability Act* requires provincial agencies to establish evaluation and reporting systems. The results-based British Columbia Forest Practices Code will include establishment of monitoring and evaluation systems to determine whether or not desired results have been reached. “Results-based” and “accountability” have become routine words used by leaders of many government and government-funded programs in British Columbia and elsewhere.

Organizations also conduct program evaluation to improve programs, to make decisions about the future of programs, and to identify cause-and-effect relationships between their activities and the outcomes of their work.

The natural resource sector does not have a long history of using evaluations. Many natural resource organizations lack evaluation capacity. For many people, program evaluation is still a mysterious notion. This publication is designed to “de-mystify” program evaluation and give program leaders the “basics” to enable them to talk about evaluation and to conduct basic—but meaningful—evaluations.

## **WHO SHOULD USE THIS PUBLICATION?**

---

This publication is intended for natural resource professionals, and program and project managers. Examples could include Community Forest pilot program leaders, Innovative Forest Practices Agreement (IFPA) holders, government-funded researchers, educators, forest practices staff, and many others.

## **WHAT IS EVALUATION?**

---

In this publication, “evaluation” is a management tool that involves measuring and reporting on the results of programs and projects.

Case, Andrews, and Werner (1988) provide another definition: “... to make an explicit judgement about the worth of all or part of a program by collecting evidence to determine if acceptable standards have been met.”

Evaluation is both an art and a science. The art of evaluation involves identifying purposes and audiences, creating appropriate designs, and interpreting data about a program, project, or policy. The science of evaluation involves systematically gathering and analyzing evidence about the impacts.

When evaluations are conducted by parties outside an organization, it is usually called “auditing.” In this publication, evaluation means an *internal look*—within an organization—at how well outcomes match what was intended.

## **DETERMINING THE PURPOSE OF AN EVALUATION**

---

Each time you begin an evaluation, you should carefully identify the purpose of the evaluation and who will use the results. Evaluations may be used to:

- guide decisions about budgets,

- communicate with the public and with political decision makers,
- identify strengths and weaknesses of a program, or
- determine whether or not to repeat or continue a program.

Do not assume that all parties agree on the purpose. Discuss the purpose in explicit terms.

## **TYPES OF EVALUATIONS**

---

Michael Quinn Patton (1997) describes three primary uses of evaluation:

- to judge merit or worth of program (such is the purpose of a “summative evaluation”),
- to improve programs (such is the purpose of a “formative evaluation”), or
- to generate knowledge.

A *summative* evaluation is conducted at the end of a program to help decision makers decide a program’s future.

A *formative* evaluation is conducted during the life of a program to identify its strengths or weaknesses and enhance its quality and effectiveness.

An evaluation conducted to *generate knowledge* is usually carried out by academics examining trends or causal links across programs.

Program evaluation is related to policy evaluation, however “programs” typically include human resources and infrastructure while “policies” are more likely to be regulations or standards with or without infrastructure.

## **DOES EVALUATION INVOLVE RESEARCH?**

---

To collect reliable information about a program’s effectiveness, managers must know a few things about basic research methodology. However, evaluation uses a wide array of research methods ranging from observation to experimental research. In-depth statistics are not often necessary.

Evaluation does, however, require that managers identify “researchable questions” that can be answered by collecting and analyzing data about their program. Evaluators typically try to identify cause-and-effect relationships between an activity designed to induce change (such as a workshop) and a particular desired outcome (such as increased knowledge of participants).

Evaluation data can be obtained both quantitatively and qualitatively. The word “data” refers not only to numbers, but also to interviews, document reviews, observation, and case studies.

## **FRAMING YOUR THINKING ABOUT EVALUATION**

---

Some managers find it helpful to think of evaluation in terms of a ladder, as shown in Figure 1 (Bennett 1976). The “lowest” level of evaluation (i.e., Inputs) refers to resources expended for programs or projects. At this level you might report, “We devoted 23 person-days of staff time to this project.” This statement does not say very much about the difference that the program or project made. Ideally, you want to report from a higher level on the ladder.

Further up the ladder, you document changes in knowledge, attitudes, skills, and aspirations (or

plans) of the target audience. You might report, “Seventy percent of participants said that they plan to use the new practice in their operations.” This kind of evidence is slightly more convincing to someone judging whether your program was worthwhile than if you simply say, “We spent 23 days on the project.”

Further up the ladder yet, you report on changes in practices or behaviours of your target audience as a result of your efforts—such as, “Seventy percent of program participants incorporated the new practice in their operations.” This is even better evidence of a program’s worth, but requires you to survey participants to see if they incorporated the new practices into their operations.

The top rung of the ladder shows results related to the long-term objectives that drive the program or project—for example, “protection of biodiversity values.” Managers cannot typically evaluate at the top level, because they cannot isolate other factors that may have led to the long-term result, but it is nevertheless helpful to know what the ultimate goal is.

Every program or project manager should establish desired outcomes at the outset of a program, project, or policy. The higher up the ladder, the more time it takes to gather data about outcomes, but the more convincing the evidence will be. As a manager, you must decide the trade-off between strong evidence of worth, and cost/time required to gather evidence.

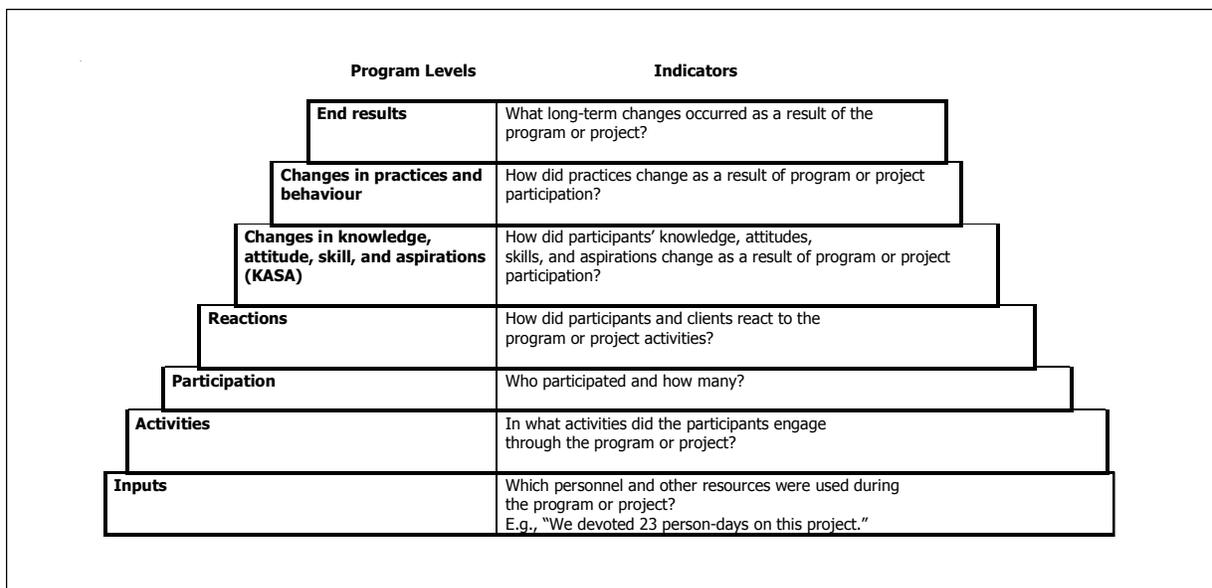


FIGURE 1. *Hierarchy of Program Evaluation Evidence.*<sup>1</sup>

## STEPS FOR CONDUCTING PROGRAM EVALUATION

*Step 1. Identify purpose.* Clearly identify the reason for conducting the evaluation. Identify the scope of the evaluation. Do you want to evaluate the whole program or just a part of it? Which component? What do you want the evaluation to tell you? Ensure everyone on the team agrees.

*Step 2. Review program goals.* Closely examine the program or project goals as stated by the designers of the program or project. What changes did the designers hope to make?

*Step 3. Identify evaluation stakeholders.* Stakeholders are those who have a stake in the outcome of the evaluation, not the audience(s) targeted by the program or project. Stakeholders include people both inside and outside the organization.

1. Adapted from: Bennett, C.F. 1976. Analyzing impacts of extension programs. US Department of Agriculture, Washington, D.C.

- Step 4. Contact stakeholders.* Obtain input about what questions they have about the program, project, or policy. This can be accomplished through a workshop session where evaluation questions are brainstormed, or by contacting stakeholders individually.
- Step 5. Revisit the purpose of the evaluation.* Based on your conversations with stakeholders and your own reason for initiating the evaluation, rewrite the purpose of the evaluation. Consider Bennett's hierarchy (Figure 1)—how far up the ladder do you need to go to demonstrate the effects of your program or project? Write your list of “evaluation questions.”
- Step 6. Decide if evaluation will be in-house or contracted out.* Based on the scope of the evaluation and the nature of the evaluation questions, decide whether you need to hire a professional evaluator or if you can conduct the evaluation with existing staff. Develop a budget based on your decision. See Section J for more information regarding external and internal evaluators.
- Step 7. Determine data-collection methods.* Decide on data-collection procedures to answer your evaluation questions. Refer to the appendix of this publication to help select methods that will best address your evaluation questions.
- Step 8. Create data-collection instrument.* Construct or adapt existing data-collection instrument(s) (such as surveys or interviews). Whether you design your own or adapt from other instruments, test the draft instrument for validity and reliability before administering it. See Appendix D for discussion about validity and reliability.
- Step 9. Test data-collection instrument.* Administer the draft instrument with a group of willing respondents and ask for their feedback on the instrument. Which questions were not clear? Were any questions misleading?
- Step 10. Collect evaluation data.* See Appendix A for more information about data-collection methods.
- Step 11. Summarize and analyze the data.* See Appendix C for more information about analyzing data and interpreting results.
- Step 12. Prepare reports for stakeholders.* See page 6, Reporting Evaluation Results.

Some inexperienced evaluators are tempted to jump quickly to Step 10 (collect evaluation data), but it is important to devote nearly as much time to the earlier steps as the later steps. In particular, avoid the temptation to skip Step 4 (contact stakeholders). You do not want to complete your evaluation only to have important decision makers say, “Yes, but . . . what about . . . .”

---

## **WHEN SHOULD AN EVALUATION BE CONDUCTED?**

Evaluation data should be collected at several (and sometimes all) stages of a program or project. These stages include:

- *The design stage.* When information is collected before a project begins, it is called a “needs assessment.” Knowing the needs of a target audience helps determine desired outcomes.
- *The start-up stage.* Evaluation information gathered at the beginning of a program or project helps establish a baseline to which changes can later be compared. This usually involves conducting a pre-test or other ways of gathering information about existing conditions.
- *While the program or project is in progress.* Collecting information during a program or project helps managers determine if adjustments are needed.
- *After the program wraps up.* A summative evaluation “sums up” what has occurred in the project, asks for end-of-project reactions, and assesses success in meeting objectives. It is typically used for accountability purposes.
- *Long after the program finishes.* This stage of evaluation looks at the long-term benefits of a program.

Before any project or program can be evaluated, it is critical that objectives and desired results of the program or project be clear. Ideally, expected outcomes will be identified before the program begins; however, in reality this is not always the case. At the very least, a program should have clear objectives by the time an evaluation is conducted so evaluators know what standards to use in judging a program, project, or policy.

Developing an “evaluation culture” within an organization means project staff routinely collect evaluation information as activities take place. This is a more desirable and effective practice than waiting until a program or project is nearly complete before thinking about evaluation.

## INTERNAL VERSUS EXTERNAL EVALUATORS

---

How do managers decide whether to use “in-house” staff to conduct evaluations or to hire external evaluation consultants? Table 1 shows the advantages and disadvantages of using internal and external evaluators.

One option is a “hybrid”—i.e., using internal staff members who contract out portions of the evaluation to professional evaluators when appropriate. This hybrid approach helps organizations develop their internal capacity. A professional can coach a team through the steps, as well as conduct certain parts of the evaluation.

FIGURE 2 *Using internal versus external evaluators.*<sup>2</sup>

Options	Advantages	Disadvantages
Using internal staff as evaluators	<ul style="list-style-type: none"> <li>– Familiar with organization</li> <li>– Credible within organization</li> <li>– Develops internal evaluation capacity</li> </ul>	<ul style="list-style-type: none"> <li>– Potential for lack of objectivity</li> <li>– Burden of additional tasks on staff</li> <li>– Potential lack of power</li> <li>– May lack evaluation skills</li> </ul>
Hiring external evaluators	<ul style="list-style-type: none"> <li>– Has specialized skills</li> <li>– Has independence and objectivity</li> <li>– Has readily available skills</li> </ul>	<ul style="list-style-type: none"> <li>– Lack of knowledge of organization</li> <li>– Limited access to information and people</li> <li>– Potential for extra expense</li> </ul>

## DEVELOPING CLEAR, SPECIFIC, AND MEASURABLE OBJECTIVES

---

Successful evaluation relies on designers developing clear, specific, and measurable objectives at the outset of a program or project. Objectives should state what changes or outcomes are expected as result of the program or project. They should begin with the statement: “As a result of our program or project, \_\_\_\_\_.” Complete the sentence by stating the intended effect or influence of the program on the targeted audience.

Here is an example of a measurable objective:

As a result of the woodlot licensee workshops, participants will be able to develop stand-management prescriptions such as surveying, selecting treatments, and drafting the prescription according to standards and guidelines.

---

2. Adapted from: Boyle, R. and D. LeMaire (eds). 1999. Building effective evaluation capacity: lessons from practice. Transaction Publishers, New Brunswick, N.J.

To find out if this objective has been achieved, you can observe or test workshop participants during the workshop.

Remember to avoid objectives that refer only to what you will do, rather than what will be different as a result of your work. For example, “Our objective is to provide 10 training sessions” is measurable but says little about the expected impact of your program.

## **AN EXAMPLE OF PROGRAM EVALUATION**

---

In 1997, the Research Branch of the British Columbia Ministry of Forests conducted a formative evaluation of biodiversity publications that had been released during the previous several years by the Royal British Columbia Museum and the provincial ministries of forests and environment.

The evaluation was designed to reveal how clients prefer to get biodiversity publications and to measure client satisfaction with the existing distribution system. The agencies hoped to use the evaluation results to improve exposure of future publications. The evaluation was led by staff members who hired a contractor to conduct the telephone surveys and write a report.

The contractor conducted a telephone survey of people selected randomly from a list of clients. Respondents were asked questions regarding familiarity, content, and distribution of the publications. Based on the findings, improvements were made in distribution of the publications.

## **REPORTING EVALUATION RESULTS**

---

Develop the report and communicate the results to stakeholders as quickly as possible after data collection. Timely reporting is important to ensure the information remains relevant.

While evaluation findings are usually documented in an evaluation report, do not limit yourself to a written report. Select from a variety of formats such as videotape, PowerPoint presentations, public meetings, newsletters, poster sessions, question-answer periods, bulletins, brochures, and web pages. Reports should be clear and candid. In choosing the format(s), consider the audience’s level of technical knowledge and the time people have to absorb the findings.

Avoid technical words that might be unfamiliar to any stakeholder. Because most readers will not read the entire report, write an executive summary highlighting procedure, major findings, conclusions, and recommendations. Include a description of the program, purpose, and procedures, a justification of the conclusions, and recommendations.

What if the evaluation reveals negative findings? For example, the program is not meeting its objectives or is being mismanaged. Use the evaluation to describe likely causes of negative outcomes. Although negative results can help managers avoid mistakes in the future, negative findings should be reported in a manner that promotes learning and improvement, rather than feelings of failure.

Negative findings should be reported in a manner that:

- is sensitive to the feelings of stakeholders;
- presents positive findings first;
- uses positive terms such as “accomplishments,” “in progress,” and “things to work on”;
- creates an atmosphere of reflection, dialogue, and positive thinking; and
- helps stakeholders think of themselves as problem solvers.

## **WHERE TO GET MORE INFORMATION**

---

### **Program evaluation niche library, FORREX—Forest Research Extension Partnership**

This collection in Victoria includes literature on evaluation methodology. FORREX can also help partner organizations with information on any aspect of their evaluations. For more information, call FORREX at 250-363-0769.

### **Program evaluation standards book**

The evaluation profession has created a set of standards in *The Program Evaluation Standards, 2<sup>nd</sup> Edition*. Published by the Joint Committee on Standards for Educational Evaluation, it can be purchased through Sage Publications ([www.sagepub.com](http://www.sagepub.com)). The book is a plain-language guide for managers who hire consultants or conduct in-house evaluations. It covers topics such as deciding whether to evaluate, budgeting for evaluation, and designing an evaluation.

### **Canadian Evaluation Society, BC Chapter**

The Canadian Evaluation Society(BC Chapter), at <http://bc.evaluationcanada.ca/>, is a place to start when contacting professional evaluators

### **American Evaluation Association**

The American Evaluation Association at [www.eval.org](http://www.eval.org) has useful publications and links.

### **University of Wisconsin Extension Service**

This Web site includes various evaluation publications in pdf format: [www.uwex.edu/ces/pubs](http://www.uwex.edu/ces/pubs)

### **Michigan State University Center for Evaluative Studies**

This centre can provide linkages, information, and technical assistance to organizations: [www.canr.msu.edu/evaluate](http://www.canr.msu.edu/evaluate)

## **GLOSSARY**

---

**Data:** Material, gathered during the course of an evaluation, that serves as the basis for information, discussion, and inference. Data can be numeric or narrative.

**Evaluation:** Systematic investigation of the worth or merit of a program, project, or policy.

**External evaluator:** An evaluator from outside the organization who has no vested interest in the evaluation findings.

**Feasibility:** The extent to which resources allow an evaluation to be conducted.

**Formative evaluation:** Evaluation designed and used to improve a project or program, especially when it is still being developed.

**Instrument:** An assessment device adopted, adapted, or constructed for the evaluation.

**Internal evaluator:** A staff member from within the organization who conducts evaluations on behalf of the organization.

**Merit:** The excellence of a program, project, or policy, as assessed by its intrinsic qualities or performance.

**Metaevaluation:** An evaluation of the evaluation process.

**Program evaluation:** An assessment of ongoing program activities. Usually this is a more comprehensive evaluation than a project evaluation.

**Project evaluation:** An assessment of time-limited activities (i.e., funded for a defined period of time to perform a specific task).

**Qualitative data:** Evidence presented in narrative form.

**Quantitative data:** Evidence represented by numbers.

**Stakeholders:** Those who have an interest or stake in the outcome of the evaluation. I.e., usually those people who will use the evaluation results for decision making. Stakeholders are not the same as the target audience of program, project, or policy.

**Summative evaluation:** Evaluation designed to present conclusions about the merit or worth of a project, and recommendations about whether it should be retained, altered, or eliminated.

**Target audience:** People at whom a program, project, or policy is aimed. For example, the target audience of an extension publication might be practitioners.

**Worth:** The value of an object in relationship to a purpose.

Most of these terms above are adapted from:

Sanders, J. (chair). 1994. The program evaluation standards, 2nd edition. Joint committee on standards for educational evaluation. Sage Publications, Inc., Thousand Oak, CA.

## REFERENCES

---

- Alkin, M. 1990. *Debates on evaluation*. Sage Publications, Inc., Thousand Oaks, CA.
- Archer, T. and J. Layman. 1991. "Focus group interview" evaluation guide sheet. Cooperative Extension Service, Ohio.
- Bennett, C.F. 1976. *Analyzing impacts of extension programs*. U.S. Department of Agriculture, Washington D.C.
- BC Ministry of Forests. May 2002. A results-based forest and range practices regime for British Columbia—discussion paper. <http://www.resultsbasedcode.ca/>
- BC Ministry of Finance and Corporate Relations. July 2000. Bill 2–2000, budget transparency and accountability act. [http://www.qp.gov.bc.ca/statreg/stat/B/00023\\_01.htm](http://www.qp.gov.bc.ca/statreg/stat/B/00023_01.htm)
- Boyle, R. and D. LeMaire (editors). 1999. *Building effective evaluation capacity: lessons from practice*. Transaction Publishers, New Brunswick, NJ.
- Case R.; M. Andrews; and W. Werner. 1988. *How can we do it? An evaluation training package for development educators*. Research and Development in Global Studies, BC.
- Dillman, D.A. 1995. *Survey methods*. Class notes in AG\*SAT Graduate Course in Program Evaluation in Adult Education and Training, University of Nebraska-Lincoln.
- Fink, A. 1995. *How to sample in surveys*. Sage Publications, Inc., Thousand Oaks, CA.
- Fraenkel, J.R. and N.E. Wallen. 1996. *How to design and evaluate research in education*. McGraw-Hill Inc., New York.
- Krueger, R.A. 1994. *Focus groups: a practical guide for applied research*, 2nd edition. Sage Publications, Inc., Thousand Oaks, CA.
- Mueller, D.J. 1986. *Measuring social attitudes*. Teachers College Press, New York.
- Neito, R. and J.L. Henderson. 1995. *Establishing validity and reliability (draft)*. Ohio State Cooperative Extension.
- Patton, M.Q. 1986. *Utilization-focused evaluation*. Sage Publications, Inc., Newbury Park, CA.
- Rossi, Peter H. and Howard E. Freeman. 1985. *Evaluation: a systematic approach (third edition)*. Sage Publications, Inc., Beverly Hills, CA.
- Sanders, J. (chair). 1994. *The program evaluation standards*, 2<sup>nd</sup> edition. Joint committee on standards for educational evaluation. Sage Publications, Inc., Thousand Oaks, CA.
- Shuley, G. 1998. *Evaluation of biodiversity publications*. Research Branch, BC Ministry of Forests, Victoria.
- Suvedi, M. 1988. *Introduction to Program Evaluation*. Program Evaluation Measurement Resource. Michigan State University. Department of Agriculture and Extension Education. 2 May 2003. <http://www.ag.ohio-state.edu/~brick/dcc1pg.htm>
- Suvedi, M; K. Heinze; and D. Ruonavaara. 1999. *Evaluation of extension programs: a training manual*. East Lansing: Program Support Systems, Michigan State University Extension.
- Wholey, J.S.; H.P. Harty; and K.E. Newcomer (editors). 1994. *Handbook of practical program evalu-*

ation. Jossey-Bass Publishers, San Francisco, CA.

Worthen, B.R. and J.R. Sanders. 1987. Educational evaluation: alternative approaches and practical guidelines. Longman, New York.

Yin, R.K. 1984. Case study research: design and methods. Applied Social Research Methods Series, Vol. 5. Sage Publications, Inc., Newbury Park, CA.

## APPENDIX A: PROGRAM EVALUATION: METHODS OF COLLECTING DATA<sup>3</sup>

---

The method of collecting evaluation data depends on:

- type of information desired,
- time availability,
- cost,
- level of expertise using the method, and
- whether the information collected will be viewed as credible, accurate, and useful by the stakeholders of the evaluation.

No one method is the “right” or only one.

In a *program evaluation* that assesses many components, you are likely to use a combination of methods to “triangulate”—that is, use more than one source of evidence. In a less complex *project evaluation*, you may be more likely to need only one source of data.

Two types of data are used in evaluation: qualitative and quantitative.

### Qualitative Methods

Qualitative methods include methods such as interviews, document review, observation, and case studies. These methods are most appropriate for:

- Open-ended questions, to bring out perspectives.
- Understanding the context in which a program or project takes place.
- Complex problems and process issues.
- Clarifying relationships between program objectives and implementation.
- Identifying unintended consequences of a program.
- Gathering descriptive information.
- Understanding operations and effects of programs.
- In-depth analysis of the effects of the program.

### Existing documents

These include reviewing memos, diaries, photographs, records, receipts, reports, meeting minutes, proceedings, newspaper articles, or editorials. These types of descriptive data provide insights into a program that cannot be obtained in any other way. The information is typically available at minimal cost.

### Participant observation

Participant observation involves an observer who participates in activities of the people he or she is studying and records activities and behaviours. Participant observation alone rarely provides enough information for a program evaluator. Usually, more detailed information must be obtained by interviewing people with knowledge about a program (known as “key informants”), or by interviewing participants themselves.

Limitations of this method:

- If the group members are aware of being observed, their behaviour may be affected.
- Different observers may not record events consistently.
- If the evaluator is involved in an activity, he or she may lose objectivity.

---

3. This section is adapted from: Suvedi, M., K. Heinze, and D. Ruonavaara. 1999. Evaluation of extension programs: a training manual. East Lansing: Program Support Systems. Michigan State University Extension.

### ***Case studies***

This method provides in-depth information on a single unit, group, or organization through personal discussion, interaction, observation, or review of existing documents. The evaluator captures the total essence of a situation. Yin (1984) describes case studies as *explanatory, descriptive, or exploratory*. With an explanatory purpose, a case study is designed to test and *explain* causal links in real-life programs where the complexity cannot be captured by a survey. A case study can be used to *describe* the context of a program. Finally, if a program has no clear set of outcomes, it can *explore* performance measures or pose hypotheses for further evaluative work.

Case studies allow for insight into relationships and feelings. The information can be used in combination with other methods, such as survey and observation, and can be tailored to a specific situation.

Limitations of this method:

- People sometimes question the evaluator's objectivity.
- Case studies require extensive amounts of data and time.
- The findings cannot be generalized to a larger population because they focus on a limited number of cases.

### ***Personal interviews***

This method consists of collecting information from one individual at a time through face-to-face or telephone contact. The major advantages are:

- The interviewer can see the respondent and hear responses.
- The interviewer can pick up nonverbal clues and ask for clarification.
- The response rate is usually very high.

Limitations of this method:

- Interviews are costly, particularly when respondents are widely dispersed.
- Interviewing requires special skills.
- It may be difficult to keep respondents on track.
- Some respondents may feel uneasy being interviewed.

Case, Andrews, and Werner (1988) offer the following guidelines for conducting personal interviews:

- Clarify the purpose of the interview with respondents. They want to know what is expected of them, how they were selected, and if they can see the results of the evaluation. Always emphasize the confidentiality and anonymity of their responses. You may want to ask respondents for permission to take notes.
- Take notes or tape record the interview. Sometimes the exact words people use may be important.
- Focus the respondent's attention on the question. If the respondent wants to talk about something else, politely but firmly refer him/her back to the question.
- Ask all questions. Check if you omitted any questions. Be flexible when unexpected problems arise.
- Do not cut respondents off in mid-answer, even if their remarks do not seem relevant.
- Respect the respondent's right to refrain from answering a question.

### ***Focus group interviews***

Historically, focus group interviews have been used in market research to assess consumer preferences and needs, but recently they have been used in program evaluation. Focus group interviews reveal information about human perceptions, feelings, opinions, and thoughts. They are relatively inexpensive, and they stimulate dialogue among participants.

According to Krueger (1994), a focus group is typically composed of 7 to 10 participants. Members are selected because they have specific characteristics in common that relate to the topic.

Focus group interviews should be conducted by a skilled interviewer. The interviewer should create a non-threatening environment that respects different perceptions and point of view, without pressuring participants to vote, plan, or reach consensus. The discussion needs to be relaxed, comfortable, and enjoyable for participants. The focus group should be conducted several times with similar types of participants to allow the evaluator to identify trends and patterns in perceptions. Careful and systematic analysis of the discussions provides clues and insights as to how a product or service is perceived.

Steps to conducting a focus group interview:

1. Consider your purpose. Why do you want to conduct focus group interviews? Who are the users of this information? Why do they want the information?
2. Identify questions to be asked in the focus group. Establish the context for each question. Arrange the questions in a logical sequence.
3. Arrange a suitable meeting place in a convenient location such as a public meeting room, restaurant, or school. The location should be easy to find, relatively free from outside distractions, and have tables and chairs arranged with participants facing each other. Arrange for a tape recorder and test it in advance.
4. Identify the audience to be interviewed. Invite them well in advance. Reconfirm their availability to participate just prior to the interview. Explain the purpose of the meeting and how they can contribute. Prepare name tags for each participant.
5. Identify a trained interviewer (and assistant, if the group consists of more than 10 people) to conduct the focus group. The interviewer should ask questions, keep the discussion flowing, and take notes.
6. Conduct the focus group interviews. The interviewer should welcome participants and restate the purpose. He or she should reassure the participants about the confidentiality and anonymity of their responses.
7. Immediately following the focus group interview, the interviewer and assistant should discuss the common themes that surfaced during the interview. They should review the tape before the next focus group interview and make notes.
8. Analyze the results by transcribing and summarizing the taped discussion. The interpretations should focus on the meaning of the findings. Are findings of value to the stakeholders? Provide a summary of the focus group interview rather than lengthy transcriptions.
9. Prepare a short report for your stakeholders.

Limitations of the method:

- Focus group interviews are easy to set up, but require skill to conduct.
- Data interpretation is tedious, and capturing the major issues without bias is often difficult.
- Results may not be relevant to the target population.

## **Quantitative Methods**

Quantitative methods are used for comparing or ranking, classifying, and generalizing results. Quantitative methods are also:

- Suitable for large-scale projects.
- Useful for judging cause and effect.

- Often accepted as more credible by those who are more accustomed to “numbers.”
- Useful for generalizing to a larger population.

### ***Existing documents***

These include previous evaluation reports, census data, research reports, and annual reports.

### ***Surveys***

Surveys are commonly used in evaluation to document opinions, attitudes, beliefs, expectations, and aspirations. They can provide information about a population by surveying only a sample of the population. Surveys are especially useful when broad information from a large population is needed.

Surveys are generally cheaper than interviews and can reach large numbers of respondents. They allow for anonymity of responses. Evaluators can also ask relatively complex questions. Surveys allow time for respondents to reflect on events and report changes and feelings. The utility of survey data can be enhanced if the information is combined with other methods, such as observation or case studies.

Surveys can be conducted using mail, telephone, or a combination, or they can be administered under a group setting such as at the end of workshops.

Mail may be the best method when:

- size of the population is large,
- visual display of questions is needed,
- educational level of respondents is high,
- respondents are dispersed in a large geographic area, and
- the budget is low.

A mail survey should be avoided if the target population may have limited reading skills, and when survey questions are open-ended.

Telephone may be the best method when:

- respondents are widely dispersed geographically,
- speed in data collection is essential,
- sample size is small, and
- cost is not a major factor.

Telephone surveys may yield a higher response rate than mail surveys. Interviewers can explain questions not understood by the respondents.

Telephone surveys should be avoided if the questions are long and complex. Telephone surveys require specialized skills and can be costly. There can be a bias in favour of those respondents with listed phone numbers and those who are more often at home. If a respondent is unfamiliar with a caller, there might be indifference and/or poor co-operation.

*What is the difference between a telephone survey and a telephone interview?* A telephone survey is a quantitative method, whereby the surveyor asks the same question of each respondent and does not ask for elaboration. A telephone interview, however, is designed to reveal qualitative information about the context of a situation. Although interviewers have a pre-set list of questions, they are free to ask spontaneous questions to probe for more details. A telephone survey provides the basis for compilation and comparison across respondents, while a telephone interview provides in-depth understanding. A telephone interview should be used for longer and more complex, open-ended questions.

### ***Pre-/post-tests and skills tests***

Pre/post tests and skills tests measure changes in knowledge, understanding, and application of knowledge of the target audience before and after a workshop or training. Pre-tests are administered at the beginning for comparing “before and after” levels of skills and knowledge. The same test is administered at the end of the workshop or training to assess the change in knowledge or skills among participants.

These tests are commonly used in training situations where there are specific skills to be transferred. They are relatively easy to administer and can usually be carried out in a group setting. Pre-tests also introduce participants to the topics that will be presented in the workshop or training.

Limitations of this method:

- People often resist tests.
- The setting may influence test results.
- If the goal of a program or project is to induce behavioural change, tests are not appropriate measures because knowledge gain is a poor indicator of behavioural change.
- It is difficult to construct a reliable and valid test.

### **Combining Qualitative and Quantitative Methods**

Combining qualitative and quantitative methods in one evaluation effort can offset perceived biases and complement strengths of different methods. When using multiple methods, care should be taken to ensure that the methods are appropriate to the evaluation questions and that resources are not stretched too thinly. Multiple methods are appropriate for:

- Understanding a complex social phenomenon.
- Allowing for greater plurality of viewpoints and interests.
- Enhancing understanding of both typical and unusual cases.
- Generating deeper and broader insights.

## APPENDIX B: SELECTING A SAMPLE IN QUANTITATIVE DATA COLLECTION<sup>4</sup>

---

Rather than surveying every person in a given population, evaluators often survey a sample of the population. Why use samples rather than a complete count? It is cheaper in terms of time, money, materials, and effort. Using statistics, results can be accurate and precise.

Population is defined as a group of persons or items having some characteristic in common (such as readers of a publication, forestry operators, and forest planners). A good sample is a miniature version of the population. A sample is representative of the population if important characteristics (e.g., age, educational level, ethnicity, income) are similarly distributed. Sampling involves selecting a smaller number of units from among the whole group concerned in such a manner that they can be used to make estimates about the whole group.

Sampling methods are usually divided into two types, random (probability) sampling, and purposeful (nonprobability) sampling.

### Random (Probability) Sampling

Random (also called probability) sampling provides a statistical basis for claiming that a sample is representative of the target population. Samples are based on random selection of units. Every member of the target population has a known probability of being included in the sample. It eliminates subjectivity in choosing a sample.

There are several types of random (probability) samples. Evaluators frequently use the following:

- *Simple random sampling.* All individuals in the population have an equal and independent chance of being selected as a member of the sample. The list of eligible units comprising a population from which to sample is called a *sampling frame*. Members of the population are selected one at a time and independently. Once they have been selected, they are not eligible for a second chance and are not returned to the pool. One can use computer-generated lists of random numbers to select the sample. A random numbers table is sometimes used with a random starting point to identify numbered subjects.
- *Systematic sampling.* All members in the population are placed on a list for random selection and every *n*th person is chosen after a random starting place is selected. Suppose you have a list of 4500 households living in a watershed for which a sample of 450 is to be selected for surveying. Dividing 4500 by 450 yields 10, indicating that you have to select one out of every 10 households. To systematically sample from the list, a random start is needed. You can toss a die to get a number, or consider the month of the year you were born. Suppose you were born in March, the 3rd month of the year. This means that the 3rd name on the list is selected first, then the 13th, 23rd, 33rd, 43rd, and so on until 450 names are selected.
- *Stratified sampling.* To ensure that certain subgroups in the population will be represented in the sample in proportion to their numbers in the population, each subgroup, called a “stratum,” is separately numbered and a random sample is selected from each stratum. A clear rationale should exist for selecting any stratum. It is more complicated than simple random sampling and using many subgroups or “strata” can lead to a large and expensive sample.
- *Cluster sampling.* The unit of sampling is not the individual but rather a naturally occurring group of individuals, such as classroom, neighbourhood, club, and so on. The clusters are randomly selected, and all members of the selected cluster are included in the sample. Cluster sampling is used in large-scale evaluations.

---

4. The section is adapted from: Suvedi, M. 1988. Introduction to Program Evaluation. Program Evaluation Measurement Resource. Michigan State University. Department of Agriculture and Extension Education. 2 May 2003. <http://www.ag.ohio-state.edu/~brick/dcc1pg.htm>

## Purposeful (Nonprobability) Sampling

Evaluators may have to choose purposeful (nonprobability) samples if accurate listings of population are not available, resources to develop a sampling frame are limited, or obtaining co-operation from potential respondents is difficult. A purposeful sample may be chosen to ensure a wide variety of people based on a number of critical characteristics are selected. Sometimes, individuals are specifically chosen to represent a specific characteristic. More frequently, evaluators choose nonprobability samples because they can be conveniently assembled. A purposeful sample does not rely on random selection of units.

The following are common purposeful or nonprobability samples:

- *Accidental sampling.* This is the weakest type of sample, but is the easiest to get. “Man-in-the-street” interviews are typical of accidental samples. The evaluator usually uses the first 5 or 10 people who happen along and are willing to talk.
- *Reputational sampling.* This involves selecting specific people to respond to a survey or to be interviewed about an issue. The choice of an individual depends on someone’s judgement of who is and who is not a “typical” representative of the population.
- *Convenience sampling.* A convenience sample consists of individuals who are available for data collection. Households living near parks or schools, or persons working in the factory or business are chosen because of convenience.
- *Snowball sampling.* This type of sampling relies on previously identified members of a group to identify other members of the population. As newly identified members name others, the sample snowballs. This technique is useful when a population listing is unavailable.

## Determining Sample Size

Several factors need to be considered when determining sample size.

- *Cost.* Work with the smallest sample that can offer adequate data.
- *The characteristics of the population.* If the population is homogenous (i.e., people possess similar characteristics), smaller samples may be required.
- *Size of the population* and the *type of analysis* to be implemented.
- The *confidence level* and the *margin of error* of findings that you desire. In general, a confidence level of not less than 95% is sufficient when drawing conclusions from the sample. The margin of error is a matter of choice. To be relatively safe in the conclusions, a margin of error not more than 5% is generally acceptable.

Sampling error is large when the sample is small. Therefore, researchers suggest that the best option is to use as large a sample as possible. Table B-1 may be used to determine sample size based on a 5% margin of error. Notice that very large population sizes require relatively small samples. Small samples require relatively large sample sizes.

TABLE B-1 *Determining sample size based on a 5% margin of error*

Population size	Sample size (+/-5%) <sup>a</sup>	Population size	Sample size (+/-5%) <sup>a</sup>
10	10	275	163
15	14	300	172
20	19	325	180
30	28	350	187
40	36	375	194
50	44	400	201
65	56	450	212
75	63	500	222
90	73	1 000	286
100	81	2 000	333
125	96	3 000	353
150	110	4 000	364
175	122	5 000	370
200	134	6 000	375
225	144	8 000	381
250	154	10 000	385
275	163	1000 000	398

<sup>a</sup> Note: These numbers represent the total number of returned surveys needed to form a valid sample, not the number of surveys mailed. Usually, surveyors must mail many more surveys than they expect to have returned.

### **Analyzing Qualitative Data**

Analyzing qualitative data requires synthesis skills, interpretative skills, and insight. Ethnographic software such as NUD\*IST is sometimes used in analyzing qualitative data. Since the data is in word form, the software helps code the results by words frequently found in the data.

### **Analyzing Quantitative Data**

Quantitative data provides numerical values to each response. It is typically collected through a pre-determined list of closed-ended questions from which respondents choose options. The quantitative data is relatively easy to tabulate and analyze. Simple descriptive statistics, including measures of variability (i.e., frequency, range, percentile, standard deviation) and measures of central tendencies (i.e., mean or average, median, mode), are used to present the findings.

Correlation coefficient is used to determine the linear relationship between variables. Cross-tabulations are used to determine associations. Evaluators use t-test to determine differences in mean scores between two groups, and analysis of variance (ANOVA) to determine differences in mean score when three or more groups are involved. Computer software such as Statistical Package for Social Sciences (SPSS), Minitab, and Mynstat are used to analyze quantitative data.

### **Interpreting the Results**

Interpretation of results involves making decisions about what the data reveals about the program or project. Sometimes evaluators collect evidence and make judgements about a project without referring back to the original standards. Judgement should be made only by comparing evidence against predetermined standards for the program, project, or policy. While making a judgement, make sure that the effects you want to attribute to the project are in fact outcomes of the project itself and are not attributable to other factors. If there are any unintended outcomes of the project worth mentioning, do so in your findings.

## APPENDIX D: ESTABLISHING VALIDITY AND RELIABILITY<sup>5</sup>

---

One of the most important steps in creating an evaluation instrument such as a survey is to establish its validity and reliability. According to Mueller (1986), validity and reliability are the most important criteria for assessing the quality of instruments.

**Validity** asks the question, “Does the instrument measure what it purports to measure?”

The following procedure is recommended to establish the validity of an instrument:

1. Clearly define what you want it to measure (e.g., reactions, knowledge level, people involvement, behaviour change, etc.).
2. Prepare a draft of your instrument. Search for existing instruments related to your topic as a guide in developing your own instrument. You may use similar question formats and response categories.
3. Recruit five to seven persons to serve as a panel of experts for reviewing your instruments in terms of content, format, and audience appropriateness. Remember that the members of the panel should be familiar about the purpose of the study. Ask the panel of experts to review the instrument and give feedback.
4. Revise the instrument by incorporating the suggestions offered by the panel of experts.
5. Field test the instrument to test its suitability and clarity. Select about 10 persons who are similar to the target audience to participate in the field test. If possible, watch people complete the questionnaire. Watch for hesitation, erasures, or skipped questions. Seek verbal feedback after you have watched them complete the instrument. If some respondents appear confused or hesitant, ask why. Based on the feedback, revise your instrument.

**Reliability** asks the question, “Does the instrument consistently yield the same results with the same group of people under the same conditions?” Reliability looks for consistency, accuracy, and dependability of an instrument. Usually, reliability is established by conducting a pilot test.

A test-retest method can also be used to establish reliability. This method involves administering the same instrument twice to the same group. According to this method, 15–20 persons having characteristics similar to the target audience (but the people are different from those in the field test group described above) are asked to complete the entire instrument. After about two weeks, the same instrument is re-administered to the same group of people. Responses on each question are compared in pairs, i.e., first-time and second-time answers from the same individual are compared. A high degree of agreement (70% or higher) between the paired scores indicates that the instrument is reliable (Neito and Henderson 1995).

“Internal consistency” methods of assessing reliability require that an evaluation instrument be administered only once. There are several internal-consistency methods for establishing reliability. Some frequently used methods are:

- *Split-half procedure*. This procedure involves scoring two halves (usually odd items versus even items) of a test separately for each person and then calculating a correlation coefficient for the two sets of scores. The coefficient indicates the degree to which the two halves of the test provide the same results, and hence describes the internal consistency of the instrument (Fraenkel and Wallen 1996).
- *Kuder-Richardson approach*. This approach utilizes three pieces of information about a test—the number of items in the test, the mean, and the standard deviation. It assumes that the items are of

---

5. This section is adapted from: Suvedi, M. 1988. Introduction to Program Evaluation. Program Evaluation Measurement Resource. Michigan State University. Department of Agriculture and Extension Education. 2 May 2003. <http://www.ag.ohio-state.edu/~brick/dcc1pg.htm>

equal difficulty. Manual computation of reliability coefficient is fairly complicated. Computer programs are available for testing various types of reliability coefficients.

- *Alpha coefficient.* Cronbach alpha is another procedure you can use to check on internal consistency of an instrument. This procedure is done by calculating reliability of items that are not scored right versus wrong. This procedure is appropriate to establish reliability of questions asked on a scale designed to measure reactions, attitudes, or perceptions (e.g., 1=strongly disagree, 2=disagree, 3=neutral/no opinion, 4=agree, 5=strongly agree).

A frequently asked question about reliability is, “What value of reliability coefficient is adequate to establish the instrument’s reliability? Is a reliability coefficient of 0.50 good or bad?” A useful rule of thumb is that reliability should be at least 0.70 and higher.

## APPENDIX E: EXAMPLE OF END-OF-WORKSHOP EVALUATION FORM

---

### Agroforestry Workshop

## PARTICIPANT FEEDBACK

Please help us improve future workshops by completing this evaluation form. Your responses will be anonymous. Circle an appropriate number for each item on the scale provided, or enter the most appropriate response.

#### A. How would you assess the impact of this workshop on:

	No change		Some improvement		Greatly improved
1. Your knowledge about agroforestry	1	2	3	4	5
2. Your attitude about agroforestry as a viable option	1	2	3	4	5
3. Opportunity to exchange ideas with other people	1	2	3	4	5
4. Your skills in incorporating agroforestry into your operations	1	2	3	4	5

#### B. Please rate this workshop based on its usefulness to you:

	Usefulness of workshop				
	Low		Moderate		High
1. Met with new people engaged in agroforestry	1	2	3	4	5
2. Exchanged ideas with colleagues	1	2	3	4	5
3. Learned about resources available to me	1	2	3	4	5

#### C. Please indicate the extent to which you agree or disagree with the following statements about this workshop:

	Agree			Disagree	
1. The topics presented in the workshop were appropriate and timely	1	2	3	4	5
2. The presentations were easy to understand	1	2	3	4	5
3. The topics presented met my expectations	1	2	3	4	5
4. My personal objectives for attending the workshop were met	1	2	3	4	5
5. The workshop was worth the time I invested	1	2	3	4	5
6. Workshop length was adequate	1	2	3	4	5

#### D. How do you feel about the workshop facility?

	Poor			Excellent	
1. Meeting rooms	1	2	3	4	5
2. Meals/coffee breaks	1	2	3	4	5

#### E. Please suggest some topics for next year's workshop:

**F. Please list any other comments and/or suggestions and/or criticisms:**

**G. Personal information:**

1. Have you attended a previous FORREX workshop?  Yes  No
2. In which part of BC do you work?  North  Southern Interior  Coast

**Thank you for your help.**

## NOTES

---