

How to Conduct Evaluation of Extension Programs

**Murari Suvedi
Kirk Heinze
Diane Ruonavaara**

**ANRECS Center for Evaluative Studies
Dept of ANR Education and Communication Systems
409 Agriculture Hall
Michigan State University Extension
East Lansing, MI 48824**

December 1999

Introduction

Evaluation in extension used to focus primarily on judging a program's merit or worth. Additionally, the methodology associated with earlier forms of evaluation was portrayed as basically a quantitative activity. In today's increasingly complex and demanding world, evaluation must deal with issues of accountability, good management, knowledge building and sharing, organizational learning and development, problem identification and policy formation. As the scope of evaluation expands, qualitative approaches and multiple methods are becoming increasingly necessary. Concurrently, today's evaluator in extension finds that he or she needs to fulfill multiple roles and be familiar with numerous methods. This manual is designed to cover the expanding field of evaluation as it applies to extension and to provide you, the evaluator, with a methodological toolbox containing a broad array of methods and suggestions as to their appropriate use.

What is Evaluation?

Program evaluation is a continual and systematic process of assessing the value or potential value of Extension programs to guide decision-making for the program's future.

When we evaluate...

- ▶ We examine the assumptions upon which an existing or proposed program is based.
- ▶ We study the goals and objectives of the program.
- ▶ We collect information about a program's inputs and outcomes.
- ▶ We compare it to some pre-set standards.
- ▶ We make a value judgment about the program.
- ▶ We report findings in a manner that facilitates their use.

Why Evaluate?

Demands on Extension for program efficiency, program effectiveness and for public accountability are increasing. Evaluation can help meet these demands in various ways.

- **Planning**

- To assess needs.
- To set priorities.
- To direct allocation of resources.
- To guide policy.

- **Analysis of program effectiveness or quality**

- To determine achievement of project objectives.
- To identify strengths and weaknesses of a program.
- To determine if the needs of beneficiaries are being met.**
- To determine the cost-effectiveness of a program.
- To assess causes of success or failure.

- **Direct decision-making**

- To improve program management and effectiveness.
- To identify and facilitate needed change.
- To continue expand or terminate a program.

- **Maintain accountability**

- To stakeholders.
- To funding sources.
- To the general public.

- **Program impact assessment**

- To discover a program's impact on individuals and/or communities.

- **Advocate**

- To gain support from policy makers and advisory councils.
- To direct attention to needs of particular stakeholder groups.

When to Evaluate

There are several basic questions to ask when deciding whether to carry out an evaluation. If the answers to these questions are “No”, this may not be the time for an evaluation.

- Is the program important or significant enough to warrant evaluation?
- Is there a legal requirement to carry out an evaluation?
- Will the results of the evaluation influence decision-making about the program? Will the evaluation answer questions posed by your stakeholders or those interested in the evaluation?
- Are sufficient funds available to carry out the evaluation?
- Is there enough time to complete the evaluation?

Role of the Evaluator

The role of an evaluator is continually expanding. The traditional role of an evaluator was a combination of expert, scientist and researcher who uncovered clear-cut cause-and-effect relationships. Today evaluators are often educators, facilitators, consultants, interpreters, mediators and/or change agents.

An Evaluator's Credibility

An evaluator is judged by his or her competence and personal style. Competence is developed through training and experience. Personal style develops over time through a combination of training, experience and personal characteristics.

Competence

- ❖ Background in the program area being evaluated.
- ❖ Capacity to understand a program's context, goals and objectives.
- ❖ Conceptual skills to design the evaluation.
- ❖ Mastery of qualitative and quantitative approaches to evaluation data collection.
- ❖ Basic quantitative and qualitative data analysis skills.
- ❖ Report writing and presentation skills.

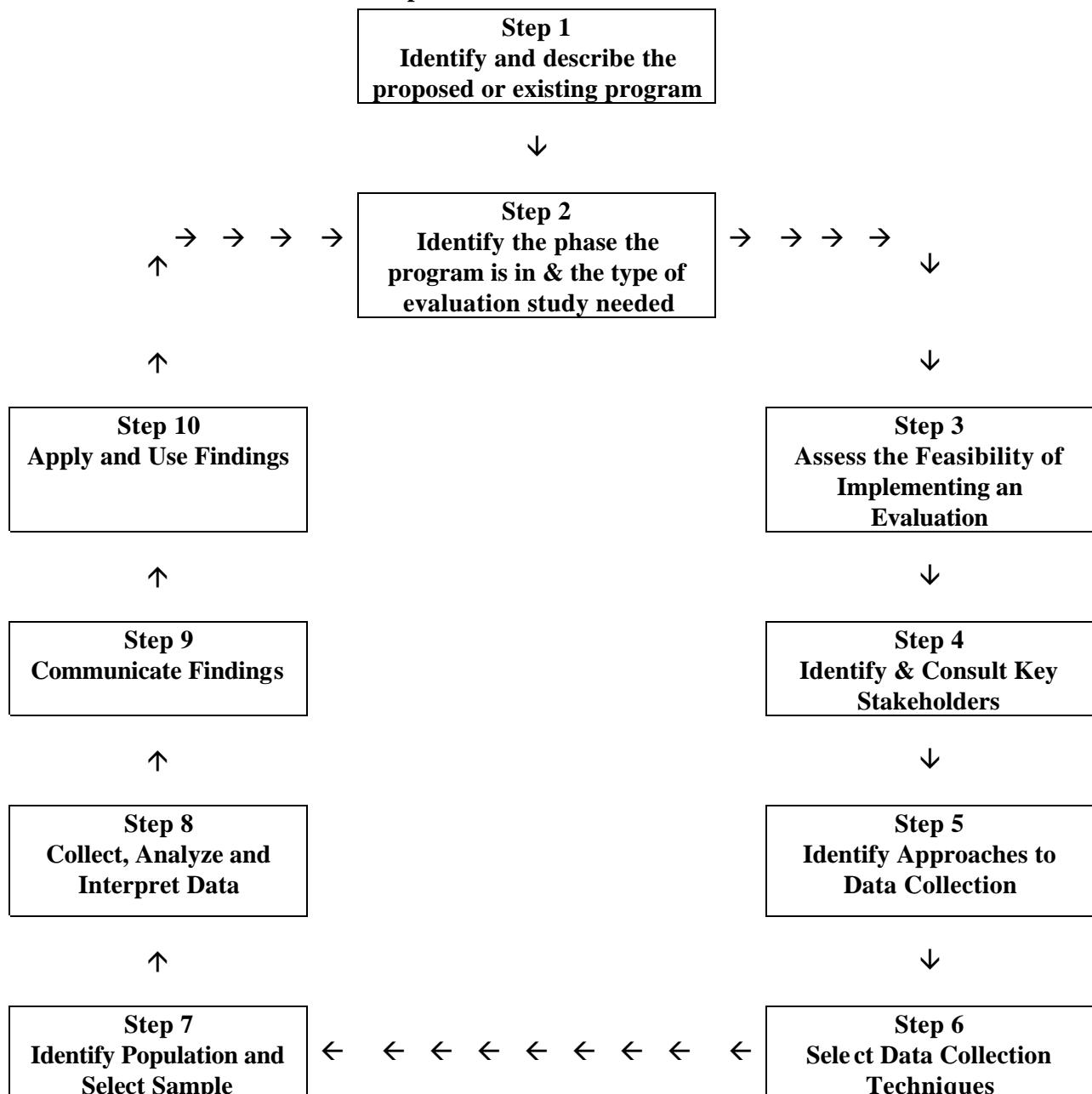
Personal Style

- ❖ Communication skills.
- ❖ Confidence.
- ❖ Strong interpersonal skills.
- ❖ Ability to nurture trust and rapport.
- ❖ Sensitivity in reporting.

Steps to Evaluation

Program evaluation can be an overwhelming process. To make program evaluation less intimidating and more manageable it can be broken down into several manageable steps. The specifics of each step may vary, depending on the nature, scope and complexity of the programs and the resources available for conducting the evaluations. These steps will be expanded upon in later sessions.

10 Steps to Evaluation – A Flow Chart



Step 1. Identify and describe the program to be evaluated

- Identify and describe the program you want to evaluate.

A description should include:

- its goals and objectives.
- the geographic boundaries of the program.
- the clientele served.
- the program funders.
- the program staff.

- Identify the audience from whom you will gather information.

Step 2: Identify the program phase & the appropriate type of evaluation study

There are a number of types of evaluation studies: needs assessments, baseline studies, formative evaluations, summative evaluations and follow-up studies. The type of evaluation study utilized is selected on the basis of stage of program, program requirements and stakeholders' interests.

Identifying the program phase and type of evaluation study needed

Ask:		Identify program phase:		Select type of evaluation study:
Is the program at a design stage?	<input type="checkbox"/>	Program design	<input type="checkbox"/>	Needs assessment
Is the program just beginning?	<input type="checkbox"/>	Program start-up	<input type="checkbox"/>	Baseline study
Is the program active?	<input type="checkbox"/>	On-going program	<input type="checkbox"/>	Formative evaluation
Is the program ending?	<input type="checkbox"/>	Program wrap-up	<input type="checkbox"/>	Summative evaluation
Is the program over?	<input type="checkbox"/>	Program follow-up	<input type="checkbox"/>	Follow-up study

Types of Evaluation Studies

A needs assessment focuses on identifying needs of the target audience, developing a rationale for a program, identifying needed inputs, determining program content, and setting program goals. A needs assessment asks questions about what exists and what is needed:

What do we need and why?
What does our audience expect from us?
What resources do we need for program implementation?

A baseline study establishes a benchmark from which to judge future program or project impact. A baseline study asks questions about what exists:

What is the current status of the program?
What is the current level of knowledge, skills, attitudes and beliefs of our audience?
What are our priority areas of intervention?
What are our existing resources?

A formative, process, or developmental evaluation provides information for program improvement, modification, and management. A formative evaluation asks descriptive questions:

What are we supposed to be doing?
What are we doing?
How can we improve?

A summative, impact, or judgmental evaluation focuses on determining overall success, effectiveness, and accountability of the program. It helps make major decisions about a program's continuation, expansion, reduction, and/or termination. A summative evaluation asks questions about what happened:

What were the outcomes?
Who participated and how?
What were the costs?

A follow-up study examines long-term effects of a program. A follow-up study asks questions about long-term impacts:

What were the impacts of our program?
What was most useful to participants?
What are the long-term effects?

Step 3. Assess the feasibility of implementing an evaluation study

Assessing the feasibility of a program evaluation helps ensure that the program can be meaningfully evaluated and that the evaluation will contribute to improving program design and/or performance. Consider the following questions carefully and then decide whether this is an appropriate time to begin a program evaluation. If the answers to many of these questions are “No”, this may not be an appropriate time to implement an evaluation study.

- Is there an important decision to be made on the basis of the evaluation?
- Is there a commitment to use the evaluation findings?
- Will important program decisions be made regardless of evaluation findings?
- Is there a legal requirement to carry out an evaluation?
- Does the program have enough impact or importance to warrant formal evaluation?
 - Is this a one-time program?
 - Will this program continue?
 - Is the cost of the program so low that an evaluation is unnecessary?
- Is it likely that the evaluation will provide valid and reliable information?
- Is it likely that the evaluation will meet acceptable standards of propriety?
 - Will the evaluation violate professional principles?
 - Is the evaluation threatened by conflict of interest?
 - Will the evaluation jeopardize the well-being of program participants?
- Is the program ready to be evaluated?
 - If a summative evaluation is suggested, has the program been operating long enough to provide clearly defined outcomes?
- Are there sufficient human and monetary resources available to carry out an evaluation?
- Is there enough time to complete the evaluation?

Step 4: Identify and consult key stakeholders

Stakeholders are people who have a stake or vested interest in the evaluation findings. They can be program funders, staff, administration, clients or program participants. It is important to clarify the purpose and procedures of an evaluation with key stakeholders before beginning. This process can help determine the type of evaluation needed and point to additional reasons for evaluation that may prove even more productive than those originally suggested.

Come to agreement with stakeholders on:

- What program will be evaluated, what it includes and excludes.
- The purpose of the evaluation.
- The goals and objectives of the program. Program goals and objectives can be written as statements indicating what the program will achieve and what criteria will be used to judge whether the objectives have been met.

Each objective should:

- contain one outcome.
- identify the target audience.
- specify what you expect to change as a result of program participation.
- be specific enough to be measurable.

Example: Members of every household in Ingham county will increase their awareness about water quality by participating in a survey conducted by Michigan State University.

- The indicators and criteria that will be used to judge value or worth of the program. When program objectives are clearly stated, the indicators and criteria to judge merit or worth will be explicitly stated.
- The questions and issues the evaluation will address.
- Who will participate in the evaluation?
- The budget and time available for the evaluation.
- The role of the evaluator.
- Who will receive the evaluation results?

Clarify evaluation questions, issues, indicators and criteria

Evaluations are conducted to answer specific questions, to address programmatic issues, to plan for future programs and/or to apply criteria to judge value or worth of an existing program. If the questions and issues that are being used are not clearly defined and the indicators and criteria that will be used to judge merit or worth are not well thought out, the evaluation may lack focus, be irrelevant, omit important areas of interest or come to unsupported conclusions.

Basic steps in selecting questions, issues, indicators and criteria

List questions, issues and criteria from all sources consulted.

Organize material into a manageable number of categories. Match level of program with indicators appropriate for that level -- remember that it is not possible for an evaluation to address all areas of interest.

Come to agreement with stakeholders on the degree of incompleteness that is acceptable, given monetary and time constraints.

Focus the scope of the evaluation to the crucial and practical.

In addition to talking with stakeholders, consider the following sources when you are clarifying the purpose of the evaluation and developing the questions, issues, indicators and criteria:

- Examine various evaluation models and available literature.
- Refer to professional standards and guidelines relating to the program area.
- Consult experts in the field.
- Use your professional judgment.

Coming to agreement on indicators

Indicators are variables. A variable is an operational representation of an attribute (quality, characteristic, property) of a system. Indicators are observable phenomena that point toward the intended and/or actual condition of situations, programs, outcomes and help gauge the performance of natural systems as well as human endeavors.

An indicator is a marker that can be observed to show that something has changed. Indicators can help people notice changes at an early stage of program's impact.

Characteristics of indicators:

- Relevant to the objective of the program to be evaluated.
- Understandable, that is to say, simple and unambiguous.
- Realizable, given logistic, time, technical or other constraints.
- Conceptually well-founded.
- Limited in number and can be updated at regular intervals.

Criteria for choosing indicators

- Is it measurable?
- Is it relevant and easy to use?
- Does it provide a representative picture?
- Is it easy to interpret and does it show trends over time?
- Is it responsive to changes?
- Does it have a reference to compare it against so that users are able to assess the significance of its values?
- Can it be measured at a reasonable cost, and can it be updated?

Bennett's Hierarchy of Evidence

Bennett's Hierarchy of Evidence provides a way of conceptualizing the relationships between program objectives and outcomes at different program levels. The hierarchy suggests the kind of information appropriate to measure to determine if an objective has been met. This will help ensure that the information you gather is appropriate for the level of the program you are evaluating.

Program Levels	Indicators
End results	Changes in participants' personal and working lives as a result of program participation.
Practice and behavior changes	Changes in participants' practices as a result of program participation.
Knowledge, attitude, skill and aspirational Changes (KASA)	Changes in participants' knowledge, attitudes, skills and aspirations as a result of program participation.
Reactions	How participants and clients reacted to the program.
Participation	Who participated and how many.
Activities	Activities that participants were engaged in through the program. The kinds of information and methods used to interact with program participants.
Inputs	The personnel and other resources used during the program.

Step 5. Approaches to Data Collection

There are two basic types of data collection: quantitative and qualitative. Quantitative data tend to focus on numerical data, while qualitative data are expressed in words.

Quantitative Methods measure a finite number of pre-specific outcomes and are appropriate for judging effects, attributing cause, comparing or ranking, classifying and generalizing results. Quantitative Methods are:

- Suitable for large-scale projects.
- Useful for judging cause and effect.
- Accepted as credible.
- Applicable for generalizing to a larger population.

Quantitative methods commonly used in evaluation of extension programs include, but are not limited to:

Existing information	Testing information & knowledge
Surveys	Benefit/cost analysis
Group-administered questionnaire	Personal interviews

Qualitative Methods take many forms including rich descriptions of people, places, and conversations and behavior. The open-ended nature of qualitative methods allows the person being interviewed to answer questions from his or her own perspective. Qualitative Methods are appropriate for:

- Understanding the context in which a program takes place.
- Complex problems and process issues.
- Clarify 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 program impacts.

Qualitative methods commonly used in evaluation of extension programs include, but are not limited to:

Existing Information	Personal Interview
Focus Group	Rapid Rural Appraisal
Participant Observation	Case Study
Group Interview	

Multiple Methods combine qualitative and quantitative methods within one evaluation study. This combination can be used to offset biases and complement strengths of different methods. When using multiple methods, care should be taken to ensure that the selected methods are appropriate to the evaluation questions and that resources are not stretched too thinly. Multiple Methods are appropriate for:

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

An Example of Multiple Methods: Garden Project Evaluation

In culturally and politically complex situations multiple methods are particularly appropriate. The following methods were combined in an evaluation of garden projects with indigenous and immigrant groups in the Petén of Guatemala.

Introduction to communities	Garden visits and biotic survey <ul style="list-style-type: none">• rapid rural appraisal• community maps <ul style="list-style-type: none">• participant observation• photography• mapping
Sampling strategy developed	<ul style="list-style-type: none">• botanical tour• chain and opportunistic interviews• identify key informants• gain access to key informants <ul style="list-style-type: none">• plant identification
	Focus group interviews
Focused unstructured interviews	Data analysis <ul style="list-style-type: none">• qualitative analysis• quantitative analysis of plant data
Visits to garden projects	<ul style="list-style-type: none">• six community visits• interviews with extension workers

Quality of Evidence

The validity and reliability of the data collection instrument determine the quality of evidence for quantitative methods.

Validity - The data collection instrument measures what it is supposed to measure and data collected are relevant to the specific situation or audience.

- Clearly define what is supposed to be measured.
- Locate or develop items to include in your instrument.
- Prepare a rough draft.
- Choose a “panel of experts” to review the instrument for content, format, and audience appropriateness.
- Revise instrument based on suggestions of experts.
- Field test for clarity, content, wording, and length.
- Revise instrument if necessary.

Reliability - The data collection instrument measures consistently, yielding the same results with the same groups of people under the same conditions.

- Carry out a pilot test with a small group of people who have characteristics similar to those of your target audience.
- Re-administer the same instrument to the same group a week later and compare the results. Computer soft wares are available to conduct reliability tests.
- Revise questions or items that produce inconsistent results. You may need reword some questions, add some items or delete certain questions to enhance reliability.

Step 6: Selecting Data Collection Techniques

There is no one best method to use when collecting data for project evaluation. Selection of a method or methods should be influenced by the type of information needed, the time available, and cost. Last, but not least you should consider whether the information collected will be viewed as credible, accurate and useful by your organization.

A large array of methods exist which can be used in evaluation. We will cover the following:

Quantitative Methods	Qualitative Methods
Existing Information	Focus Group
Testing Information and Knowledge	Rapid Rural Appraisal
Telephone Surveys	Case Study
Mail Surveys	Semi-structured Interviews
Group-administered Questionnaire	Participant Observation

Existing Information

Before you start to collect data, check to see what information already exists. Pre-existing information can be found in documents, reports, program records, historical accounts, minutes of meetings, letters, photographs, census data and surveys.

Existing information is useful for:

- Establishing the need for a program – use census data, media feature stories, maps, or service and business statistics.
- Describing how the program was carried out and who it reached – use program documents, log books minutes from meetings, enrollment records, media releases.
- Assessing results – use public records, local employment statistics, agency data, and evaluation of similar programs.

Advantages of using existing information:

- In most cases, it is readily available.
- It can be obtained with minimal cost and effort.
- Data with a wide variety of characteristics are available.
- It can be accessed on a continuing basis.
- It can have high credibility.

Disadvantages of existing information as a data source:

- Data tend to be descriptive and may require the evaluator to sort, discriminate and correlate.
- Some figures may represent estimates or projections rather than actual accounts.
- It may not reveal values, reasons or beliefs underlying current trends.
- Local community data are frequently limited and not always current.
- It may present a biased view of reality.

Testing Information and Knowledge

Tests can be used as a tool to measure the level of knowledge, understanding and ability that an individual possesses related to a particular program.

Advantages of using testing information and knowledge:

- It can provide an indication of knowledge level and other changes related to a particular program.
- It is relatively easy to implement.
- It can be carried out in a group setting.
- It tends to be low-cost.

Disadvantages of using testing information and knowledge:

- Adults often resist attempts to test their knowledge.
- Knowledge gain may be unrelated to behavior changes.
- Valid and reliable tests require special skills and time to develop.
- It is not appropriate for less literate audiences.

Basic Steps in Testing Changes in Knowledge and Information

Step 1. Construct a test consisting of questions that focus on the specific subject matter presented during the program.

Step 2. Phrase questions in a non-critical manner.

Step 3. Give the test to program participants either orally or in written form.

Step 4. If responses are given in a group discussion, be sure to elicit the full range of responses.

Step 5. Individual responses can be in a written format or discussed informally.

Step 6. Grade the responses or compare them against answers that have been identified as correct.

Step 7. Summarize and analyze the degree of comprehension achieved through the program.

Surveys

Surveys are a very popular method of collecting evaluation data and require a carefully designed questionnaire administered by mail, telephone or personal interviews. Surveys can be used to collect data on a participant's knowledge, attitudes, skills and aspirations, adoption of practices, and program benefits and impacts. It is the responsibility of the evaluator to ensure that ethical standards are maintained. This means that participation is voluntary and survey results are made public in a way that maintains confidentiality.

Advantages of using surveys:

- It permits fairly complex questions.
- It allows for anonymity of respondents.
- Cost is moderate.
- It is easy to reach a large number of people.
- Surveys are useful when the population is widely dispersed.

Disadvantages of using surveys:

- A survey does not easily prove a cause-and-effect relationship.
- Surveys are difficult to use in cross-cultural settings.
- Using surveys requires a fairly literate population.
- It can be difficult to find an accurate and up-to-date list of potential respondents.

Key questions that need to be answered before carrying out a survey:

- How many people are required for a valid survey?
- Who should answer the questionnaire -- a few or only a select group?
- How should the sample be selected?
- How high should the response rate be?
- How accurate will the results be?

When choosing a survey method, consider the resources you have available:

- Paid or unpaid people to carry out the survey.
- Your time frame and budget.
- A person experienced in survey work available to assist you.
- Available facilities such as telephone access.

Telephone Survey

A telephone survey consists of a written questionnaire that is read to a selected group of people over the telephone. The survey sample is often selected from a telephone directory or other lists. People on the list are interviewed one at a time over the phone.

Advantages of telephone surveys:

- They can be used when respondents are widely dispersed geographically.
- They tend to have a high response rate.
- They can address more complex questions than mail questionnaires.
- They provide a quick and efficient source of data.
- Selection of a specific respondent is easier to control than with a mail survey.
- The interviewer can explain questions to respondents.

Disadvantages of telephone surveys:

- Questions must be clear and concise.
- Surveys can be time consuming.
- Cost per response is comparable to mail surveys.
- They require interviewing skills and a trained supervisor.
- Bias may result as households without telephones or with unlisted numbers are excluded.
- Timing of calls is critical and may introduce bias.
- The telephone interviewer's voice or mannerisms may introduce bias.
- They require a reliable telephone system and an adequate location for conducting interviews.

Implementing a Telephone Survey

- Step 1. Find suitable facilities and equipment necessary to implement the survey.
- Step 2. Decide on a sampling design, including the method of respondent selection within a sampling unit. Choose the method to generate a pool of telephone numbers that will be used in sampling.
- Step 3. Prepare survey material: An advance letter if names and addresses are available; the questionnaire; a cover sheet to record identification number and a call-sheet; help sheets for the interviewer.
- Step 4. Train interviewers: background information about the survey; basics of telephone interviewing; how to use equipment; and how to fill out questionnaires and call-sheets.
- Step 5. Develop an interview schedule. Assess when you will be likely to contact respondents, during working or non-working hours. For most surveys, approximately 50 minutes is sufficient to complete an interview. Decide how to handle refusals.
- Step 6. Make calls. Decide on the number of calls to make to each number. In local surveys six to seven calls are customary. Make callbacks

Sample call sheet for telephone interviews

A call-sheet is used for each number chosen from the sampling frame. The interviewer records information that allows the supervisor to decide what to do with each number that has been processed. Call sheets are attached to questionnaires after an interview is completed.

Telephone interview call sheet				
Survey title _____				
Questionnaire identification number _____				
Area code & number () _____ - _____				
Contact attempts	Date	Time	Result code & comments	Interviewer I.D.
1				
2				
3				
4				
5				
6				
<u>Additional comments:</u>				
Result Codes				
Code	Explanation			
01	No answer after seven rings			
02	Busy, after one immediate redial			
03	Answering machine (residence)			
04	Household language barrier			
05	Answered by nonresident			
06	Household refusal			
07	Disconnected or other non-working number			
08	Temporarily disconnected			
09	Business or other non-residence			
10	No one meeting eligibility requirement			
11	Contact only			
12	Selected respondent temporarily unavailable			
13	Selected respondent unavailable during field period			
14	Selected respondent unavailable because of physical/mental handicap			
15	Language barrier with selected respondent			
16	Refusal by selected respondent			
17	Partial interview			
18	Respondent contacted - completed interview			
19	Other			

Sample Help Sheet for Interviewers

1. Name of sponsoring agency:

2. Purpose of study:

3. Contact person for survey:

4. Size of survey:

5. Identity of interviewer:

6. How respondents name was obtained:

7. Issues of confidentiality:

8. How to get a copy of results:

9. How will results be used:

Mail Survey

A mail survey is the most frequently used type of survey in evaluation of Extension programs and requires the least resources.

Advantages of using a mail survey:

- Can be used with a large sample size.
- Can be used with a widely dispersed population or one that is not accessible by telephone or personal interviewing.
- Provides a visual display of questions.
- Is free of interviewer bias.
- Enables respondents to give thoughtful answers and control the pace and sequence of response.
- Are relatively inexpensive.

Disadvantages of using a mail survey:

- The questionnaire must be short and carefully designed.
- The response rate is highly dependent on the number of contacts made with the respondent and the timing of the mailing.
- There is little control over the completeness of the response.
- Those who reply may not be representative of the target population.
- Pretesting of the questionnaire is necessary to avoid costly mistakes.
- A mail survey requires from four to six weeks to collect data.
- It requires a literate population and a reliable postal system.

Basic Steps in Implementing a Mail Survey

- Step 1. Prepare survey material. Design a written questionnaire, using an identification number on the cover of each questionnaire to track returns.
- Step 2. Pretest instrument to assure validity and reliability.
- Step 3. Select sample population
- Step 4. Develop a mailing schedule: a) Two weeks before mailing the survey, send an advance letter; b) mail the questionnaire including a cover letter and a stamped, self-addressed envelope; c) send a postcard a week or so later, thanking those who responded and reminding those who did not to return their surveys; d) three weeks after mailing the first questionnaire, send a follow-up letter stating that a response has not been received, including a replacement questionnaire and a stamped, self-addressed envelope. In developing the mailing schedule avoid holidays, especially the month of December. For most purposes, a 60 to 90 percent return rate is considered satisfactory.

Personal Surveys

Personal or face-to-face surveys are conducted by talking individually to respondents and systematically recording their answers to each question.

Advantages of a personal survey:

- It can be used with a highly dispersed population.
- It is suited for populations where a representative sample cannot be drawn.
- It can be used where there is a low literacy rate.
- There is a high degree of control over who answers the survey.
- The interviewer can increase the willingness of respondents to answer questions.
- Visual aids can be used to facilitate understanding of survey questions.
- Questions can be fairly complex.

Disadvantages of a personal survey:

- It can be expensive and time-consuming.
- Interviewers must be carefully selected and receive adequate training.
- It requires a good supervisor.
- It requires more material than a telephone or mail survey.

Basic Steps in Implementing a Personal Survey

- Step 1. Develop survey material: a) an advance letter if names and addresses are available; b) name tags for interviewers; c) an introductory letter explaining the purpose of the survey to be left with the respondent; d) an interviewer's instruction manual; e) sampling information for interviewers; and f) the questionnaire.
- Step 2. Identify and train a staff of interviewers.
- Step 3. Mail letters describing the procedure and telling residents to expect a visit from an interviewer.
- Step 4. Notify public officials about the survey.
- Step 5. Conduct interviews. A supervisor should be available by telephone while the survey is being carried out to handle any problems that may arise.
- Step 6. The supervisor should meet regularly with interviewers to edit questionnaires and answer any questions interviewers may have. Costly errors, misunderstandings, and cheating by interviewers can be detected at this time.
- Step 7. After interviews are completed, the questionnaires are returned to the survey supervisor.

General Procedures for a Survey Interview

Minimizing interviewer bias:

- Maintain a neat appearance.
- Follow the sampling plan to locate respondents.
- Be considerate and honest with the respondent.
- Understand the purpose of the study.
- Ask questions exactly as written.
- Record responses accurately.
- Be familiar with the research instrument.
- Follow sampling instructions.
- Check work for completeness.

Initiating contact:

- Introduce yourself, show your credentials.
- Remind respondent of the notification letter he or she received a few days earlier.
- Explain the purpose of the survey.
- Assure the respondent that his/her answers will remain confidential.
- Explain how respondents were chosen.

Guidelines for interviewing:

- Choose respondents following the sampling criteria.
- Conduct the interview or select a mutually convenient time to return.
- To avoid distractions, try to conduct the interview without an audience.
- Remind participants that the interview is voluntary and their responses are confidential.
- Establish rapport by expressing appreciation of the respondent's responses and willingness to participate.
- Read questions as they appear in the questionnaire and record answers accurately.
- Do not express your opinions.
- If an answer to an open-ended question is incomplete or appears irrelevant, probe to get a clearer response.
- If a respondent refuses to answer a question, do not insist to get an answer. It may jeopardize the entire interview.

Creating Quality Surveys by Avoiding Errors

Some surveys are more accurate than others. Accuracy means that survey results closely represent the population from which the sample has been drawn.

Inaccuracy can be caused by several types of errors including coverage error, sampling error, selection error, frame error, non-response error and measurement error.

Type of Error	Cause of Error	Control of Error
Coverage error	The sampling frame does not include all elements of the population.	Redraw list from which the sample is drawn to include all elements of the population.
Sampling error	A subset or sample of all people in the population is studied instead of conducting a census.	Increase the size of the sample; Use random sampling; Purge list of duplication.
Selection error	Some sampling units have a greater chance of being chosen than others are.	Use random sampling
Frame error	List is inaccurate or some sampling units are omitted.	Use up-to-date, accurate list.
Non-response error	Subjects can't be located or fail to respond.	Compare early to late respondents. If no difference is apparent, results can be generalized to the larger population. Contact about 10% of non-respondents and gather data from them. Compare these data with the respondents. If no difference is apparent, results can be generalized to the larger population. Compare respondents to non-respondents on known characteristics. If no difference is apparent, the results can be generalized to the larger population.
Measurement error	A respondent's answer is inaccurate or imprecise or cannot be compared to any useful way to other respondent's answer. This may be caused by: unclearly stated questions; unclear instructions; or respondents giving socially correct responses, not knowing the correct information or deliberately lying.	Choose appropriate method of data collection for your evaluation. Write clear, unambiguous questions that people can and want to answer. Train your interviewers carefully. Use valid and reliable instruments.

Group-administered Questionnaire

A group-administered questionnaire is handed directly to each participant in a group at the end of a workshop, seminar or program. Respondents answer the questions individually and hand them back to the person conducting the evaluation.

Advantages of a group-administered questionnaire:

- The questions have a direct relationship to recognized goals or objectives.
- It's low in cost and easy to administer.
- It's easy for respondent to complete.
- It gives immediate feedback.
- It can sample the total population.
- It can be used as a basis for discussion.

Disadvantages of a group-administered questionnaire:

- Program participants may not be representative of the wider population, therefore findings may not be generalizable.
- Questionnaires usually cover a single topic or issue.
- It takes time away from the program.

Basic steps in doing a group-administered questionnaire

- Step 1 The questionnaire is prepared following the guidelines for constructing a survey instrument. However, the objectives and instructions for completing the questionnaire are explained to the participants by the instructor, supervisor or agent. He or she also should assure participants that anonymity will be maintained.
- Step 2 The questionnaire is distributed to each participant to be filled out individually.
- Step 3 The questionnaire is collected and checked for completeness

Questionnaire Design

The overall aim questionnaire design is to solicit quality participation. Response quality depends on the trust the respondent feels for the survey, the topic, the interviewer and the manner in which the questions are worded and arranged. Consider whether the questionnaire is going to be mailed, given directly to respondents, used in a telephone survey or used in personal interviews. Before you begin it is essential to know what kind of evidence you need for the evaluation and how the information will be used.

Before you begin...

- Make a list of what you want to know and how the information will be used.
- Check to make sure the information is not already available somewhere else.
- Eliminate all but essential questions.
- As you write questions try to view them through the eyes of the respondents.

Writing the questionnaire

1. The title of the questionnaire should appeal to the respondents.
2. The type used should be large and easy to read.
3. The questionnaire should appear professional and easy to answer.
4. The introduction should identify the audience and the purpose of the survey and give directions on how to complete the questionnaire.
5. Questions should not appear crowded.
6. Each question should be numbered and sub-parts of a question should be lettered.
7. Questions should be arranged in a logical order with general questions preceding more specific ones. Easy-to-answer questions come first, followed by increasingly complex, thought-provoking, or sensitive questions. Personal or potentially threatening questions should be placed at the end.
8. Be explicit about what is required to answer each question.
9. Sufficient space should be left for answering open-ended questions.
10. Clearly indicate where branching occurs and where general questions resume.
11. Key words should be boldface or capitalize to avoid the possibility that they are misread.
12. Request for demographic information should be included near the end of the questionnaire.
13. For mail surveys, remind respondents to return the questionnaire and provide an addressed, postage-paid envelope.
14. The questionnaire should end with a "Thank You."

Special Questionnaire Design Considerations

Telephone questionnaires: Telephone questionnaires dependent on oral communication, so special attention must be paid to designing a questionnaire that will assist the interviewer as much as possible in holding the respondent's attention. Design and construction of the questionnaire are based on utility rather than aesthetics.

Introduction: Special attention is paid to the introduction because it is at this point that most refusals occur. The introduction should include:

- The interviewer's name.
- The name of the institution and city from which he/she is calling.
- How the phone number was obtained.
- A conservative estimate on how long the interview will take.
- A callback sheet for the interviewer to answer respondents' questions such as, explanation of the survey, use of the survey, why a male or female is selected, a phone number and contact to call to verify the legitimacy of the survey.

Mailed questionnaires: The appearance of a mailed questionnaire is of utmost importance. A mailed questionnaire must "sell" itself to the respondent to be returned. Therefore, considerable care should be taken in designing the format of the questionnaire.

- A simple booklet can be constructed by folding an 8 ½ by 11-inch paper in half.
- Make questions fit the page so that the respondent does not need to turn the page to answer a question.
- Provide easy-to-follow directions on how to answer the questions
- Arrange questions and answers in a vertical flow. Put answer choices under rather than beside the questions so that respondents move down the page rather than from side to side.

Designing a Questionnaire Cover Letter

1st paragraph:

- Explains the purpose of the study.
- Describes who will be answering the questionnaire.
- Assures confidentiality of responses.

2nd paragraph:

- Assures the respondent the study is useful.
- Lets the respondent know he or she is important to the success of the study.

3rd paragraph:

- Provides directions on how and when to return the questionnaire
- Explains the questionnaire identification number for facilitating follow-up.

4th paragraph:

- Reemphasizes the study's social usefulness.
- Promises a copy of survey results if desired.
- Indicates a willingness to answer any questions.
- Includes a statement of thanks, a closing and the sender's name and title.

Writing Questions

The questions used in a questionnaire are the basic components that determine the effectiveness of your survey. Writing good questions is not easy and usually takes more than one try. Consider what information to include, how to structure the questions and whether people can answer the questions accurately. Good survey questions are focused, clear, and to the point.

Every question should focus on a single, specific issue or topic.

Poor: *Which brand do you like best?*

Better: *Which of these brands are you most likely to buy?*

The objective of these questions is to measure consumer preference. The first question lacks focus, consumers may like a particular brand, but may not buy it because of its high price.

The meaning of the question must be completely clear to all respondents. Clarity ensures that everyone interprets the question the same.

Poor: *When was the last time you went to the doctor for a physical examination on your own or because you had to?*

Better: *How many months ago was your last physical examination?*

The first question could be interpreted in weeks, months, years, or by date.

Keep questions as short as possible. Short questions are easier to answer and less subject to error by interviewers and respondents. Long questions are more likely to lack focus and clarity.

Poor: *Can you tell me how many children you have, whether they're boys or girls, and how old they are?*

Better: *What is the age and sex of each of your children?*

A respondent may answer the first question ambiguously. For example, “I have two boys and a girl. They are 5, 7, and 10 years old.” It is not possible to determine the ages of each child from this response.

Questions should be written to avoid bias.

Poor: *Is it true that our agents always work long hours?*

Better: *On average, how many hours do extension agents work in their job?*

Types of Information

Questions can be formulated to elicit four types of information: 1) knowledge, 2) beliefs, attitudes and opinions, 3) behavior and 4) attributes. Any one or a combination of these types can be included in a questionnaire.

Knowledge questions include what people know and how well they understand something

What is the major cause of accidental deaths among children inside the home?

Beliefs, attitudes and opinions include people's perceptions, their thoughts, their feelings, their judgments or their ways of thinking.

Should the Clearwater Regional Education Center in Minor County continue to offer college-level and/or continuing education courses and programs?

Behavioral questions ask people about what they have done in the past, what they do now or what they plan to do in the future.

Have you or your family ever taken classes at the Clearwater Regional Education Center in Minor County?

Attributes are a person's personal characteristics, such as age, education, occupation or income. Attribute questions ask respondents who they are, not what they do.

Where do you currently live?

How many children do you have?

What percentage of your household income comes from off-farm employment?

Types of Questions

There are basically two distinct type of questions asked in a survey – closed-ended questions and open-ended questions.

Closed-Ended questions have pre-determined categories of responses from which the respondent can choose. When asking closed-ended questions make sure that all alternative response categories have been included.

Advantages of closed-ended questions:

- Are easy to answer.
- Do not require a lot of time.
- Require less interviewer training.
- Reduce interviewer bias.
- Can be used to compare and quantify individual responses.
- Facilitate data entry and analysis.

Disadvantages of closed-ended questions:

- Response categories must be known before questions can be written.
- Response categories may not be inclusive of all potential/likely responses.
- Response categories may be superficial or biased.
- Response categories may be interpreted differently by different respondents.
- Response is difficult if the response list is too long.
- They force participants to make choices.

Examples of Closed-ended Questions

1. Have you or members of your family ever taken classes at the Regional Education Center in this county? _____ Yes _____ No
2. To what extent do you agree or disagree with the new zoning code?
 1. Strongly disagree
 2. Mildly disagree
 3. Neither agree or disagree
 4. Strongly agree

Open-ended Questions

Open-ended questions allow respondents to answer in their own words rather than select from predetermined answers.

Advantages of open-ended questions:

- Stimulate free thought.
- Provide a chance for respondents to express feelings and opinions.
- Allow respondents to express themselves in their own terms.
- Can provide vignettes and material for explanatory quotes.
- Are excellent for exploratory studies.
- Can provide material to develop close-ended questions.

Disadvantages of open-ended questions:

- Difficult to analyze.
- Require more time to answer.
- Depend on respondent recall.
- Require greater interviewing skill.
- Lack response categories to help clarify questions.
- Handwritten responses may be illegible.

Examples of Open-ended Questions

1. How do you plan to use the information acquired during this training?
2. What do you think should be done to improve the 4-H program in this county?

Pre-testing Evaluation Instruments

Pre-testing is usually associated with quantitative methods though qualitative and participatory methods can be pre-tested as well albeit using a slightly different format. Pre-testing entails trying out evaluation techniques and instruments before beginning the evaluation process and to avoid costly errors and wasted effort. When possible, pre-testing should be done in circumstances similar to those anticipated during the evaluation itself. If feasible, use the same sampling plan you will use during the evaluation to select a mini-sample.

In pre-testing, we ask:

- *Are the issues to be discussed, the questions to be asked and/or the words to be used clear and unambiguous?*
- *Is the technique or instrument appropriate for the people being interviewed or observed?*
- *Are instructions for the interviewer or observer easy to follow?*
- *Are the techniques and/or forms for recording information clear and easy to use?*
- *Are procedures standardized?*
- *Will the technique or instrument provide the necessary information?*
- *Does the technique or instrument provide reliable and valid information using the criteria of the chosen data collection approach?*

You may find that you have to modify the technique or instrument after field testing it. If extensive revisions are made, a second field test may be necessary.

Focus Group

A focus group is a small group, typically 8 to 12 people who are relatively homogeneous, which is selected to discuss a specific topic in a non-threatening atmosphere. The focus group is moderated and recorded by a skilled interviewer. A focus group measures community needs and issues; citizens' attitudes, perceptions and opinions on specific topics; and impacts of a particular program on individuals and communities.

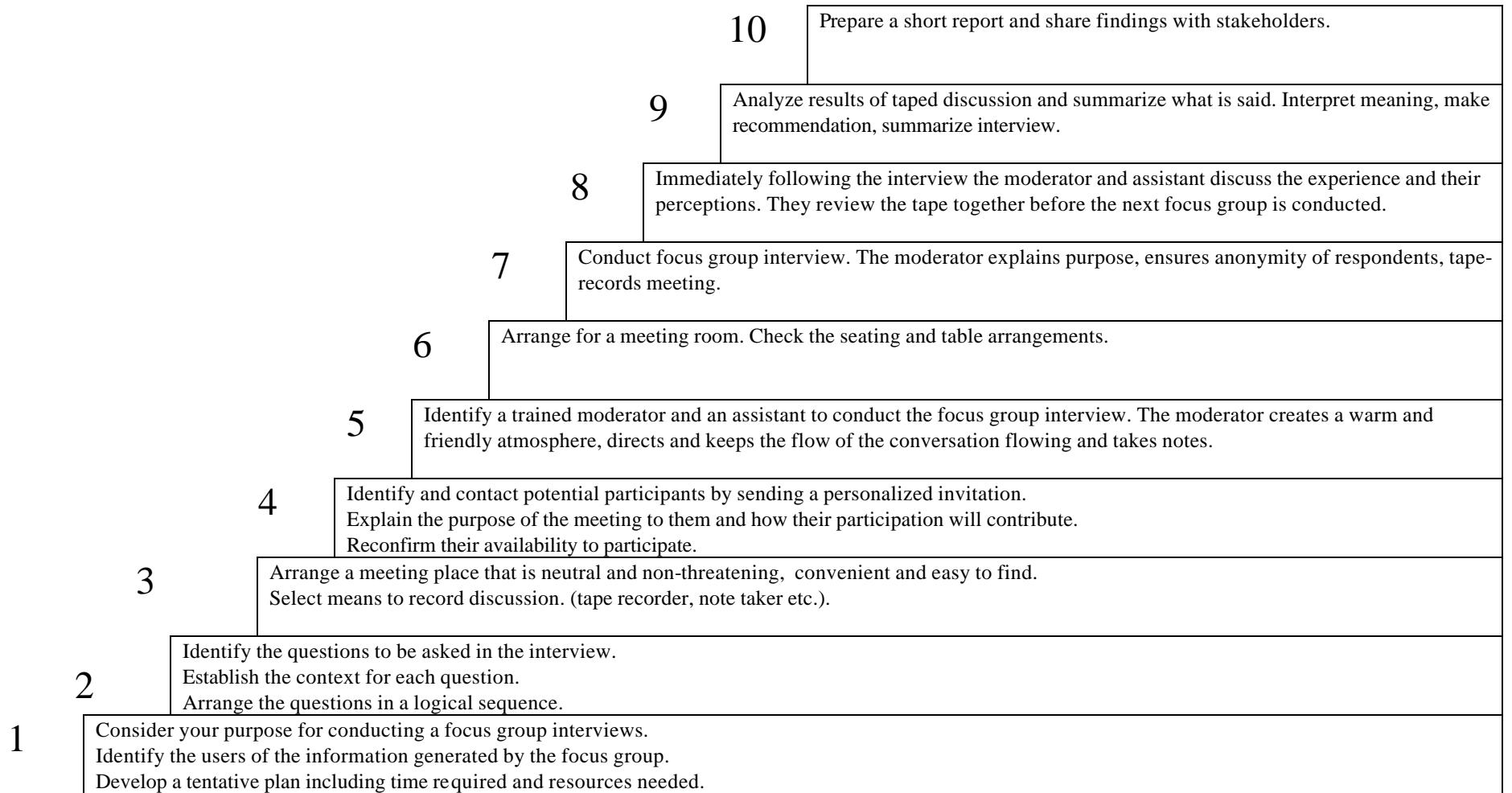
Advantages of a focus group:

- It is easy to setup.
- It is fast and relatively inexpensive.
- It can reduce the distance between project personnel and intended beneficiaries.
- It stimulates dialogue.
- It can generate ideas for evaluation questions to be included in other survey methods.

Disadvantages of a focus group:

- It is easily misused.
- It requires special moderator skills.
- Data interpretation is tedious.
- Avoiding bias can be difficult.
- Capturing major issues that emerge can be difficult.
- Results may not be generalizable to the target population.

Steps to a Focus Group Interview



How to Begin a Focus Group Discussion

The first few moments in a focus group discussion are critical. In a brief time, the moderator must create a thoughtful, permissive atmosphere, provide the ground rule, and set the tone of the discussion. Much of the success of group interviewing can be attributed to the development of this open environment. The recommended pattern for introducing the group discussion include: the welcome, the overview and topic, the ground rules and the first question.

An Example of a Typical Introduction

Good evening and welcome to our session tonight. Thank you for taking the time to join our discussion of county educational services. My name is _____ and I represent _____. Assisting me is _____ from _____. We are attempting to gain information about educational opportunities in the community. We have invited people who live in several parts of the county to share their ideas.

You were selected because you have certain things in common that are of particular interest to us. You are all employed outside the home and you live in the suburban areas of the county. We are particularly interested in your views because you are representative of others in the county.

Tonight we will be discussing non-formal educational issues in the community. These include all the ways you gain new information about areas of interest to you. There are no right or wrong answers but rather differing points of view. Please feel free to share your point of view even if it differs from what others have said.

Before we begin, let me remind you of some ground rules. Please speak up, but only one person should talk at a time. We're tape-recording the session because we don't want to miss any of your comments. If several are talking at the same time, the tape will get garbled and we'll miss your comments. We will be on a first-name basis tonight, and in our later reports, there will not be any names attached to comments. You may be assured of complete anonymity of responses. Keep in mind that we're just as interested in negative comments as positive comments, and at times the negative comments are the most helpful.

Our session will last about an hour and we will not be taking a formal break. Well, let's begin. Let's find out some more about one another by going around the room one at a time. Tell us your name and where you live.

How to ask Questions in a Focus Group

1. Use open-ended questions to stimulate discussion.

What did you think of the program?
Where do you get new information?
What do you like best about the proposed program?

2. Avoid dichotomous questions – those that can be answered with a yes or a no.

3. “Why” questions are rarely asked

“Why” questions can make people defensive and feel the need to provide an answer.
When you ask “why,” people usually respond with attributes or influences.
It’s better to ask, “What prompted you?” or “What features did you like?”

4. Use “think back” questions that remind respondents of an experience rather than ask them to speculate on the future.

5. Carefully prepare focus questions

Identify potential questions.

Five types of questions are:

- a. Opening questions (round-robin).
- b. Introductory questions.
- c. Transition questions.
- d. Key questions.
- e. Ending questions.

6. Ask uncued questions first, cued questions second

(Cues are the hints or prompts that help participants recall specific features or details.)

7. Consider using standardized questions (explain this - what does this mean?)

8. Focus the questions by using a sequence that proceeds from general questions to those focusing on specific

Sample Focus Group Questions

[Field Crops Industry Advisory Committee]

Note: These questions will be distributed to all advisory committee members during the focus group session.

1. Let's find out some more about one another by going around the room. Tell us your name, where you live and what first comes to mind when you hear the words "Michigan State University."

(Facilitator's notes: Record key words and return to some of these later)

2. What are you hearing people say about Extension agriculture and natural resources programs in your community? How has Extension's work changed in the recent past?

(Facilitator's notes: Probe positive and negative comments.)

3. Think back to an experience you have had with MSU Extension that was outstanding. Describe it.
4. Think back to an experience you have had with MSU Extension that was disappointing. Describe it. How could Extension change its programming?
5. Michigan State University Extension has adopted an Area of Expertise (AOE) team approach to Extension work. Have you taken advantage of the Area of Expertise (AOE) teams? What have been your impressions of the AOE team performance during the past year?
6. How can MSU Extension's field crop AOE team improve its future program offerings? Could you suggest ways to improve Extension field crop programs?

(Facilitator's notes: Encourage each participant to respond. Refrain from probes until each participant has a chance to react.)

7. Do you have any final comments, recommendations or thoughts about either Michigan State University Extension or its agriculture and natural resources programs?
8. Have we missed anything?

Rapid Rural Appraisal

Rapid rural appraisal (RRA) is a research approach that involves multiple data collection techniques that are quick, flexible and adaptive, yet relevant. RRA helps us learn about local people's situations, experiences and problems from a local perspective.

Advantages of rapid rural appraisal:

- Is low-cost.
- Requires little time.
- Can encourage local participation.
- Can decrease outsider bias.
- Can encourage participation of frequently overlooked groups.
- Offers flexibility in method selection.

Disadvantages of rapid rural appraisal:

- Seasonal bias.
- Accessibility bias.
- Elite bias.
- Hypothesis confirming - selective attentiveness.
- Concreteness bias - confusing specificity with generality.
- Consistency bias - premature formation of coherence in data.
- May not be generalizable.

RRA Methods Tool Box

Existing information	Visualization techniques
Individual Interviews	<ul style="list-style-type: none">• Activity mapping
<ul style="list-style-type: none">• Key informants• Oral histories	<ul style="list-style-type: none">• Time series maps (e.g., crop calendar)• Resource mapping
Group interviews	<ul style="list-style-type: none">• Social organizational mapping
<ul style="list-style-type: none">• Focus groups	<p>Ranking games</p> <ul style="list-style-type: none">• Wealth ranking• Preference ranking
	Matrices

Basic Steps to Rapid Rural Appraisal

- Step 1. Identify goals of RRA and questions to ask.
- Step 2. Identify resources available for RRA – time, skills, staff, clientele, formal organizations, informal groups, firms, transportation, telephone/mail, media.
- Step 3. Review existing documentation.
- Step 4. Identify data needs, type of analysis needed.
- Step 5. Identify possible sources of information.
- Step 6. Identify, adapt and/or create data collection methods.
- Step 7. Field test methods.
- Step 8. Adjust questions, sources of information and approaches.
- Step 9. Plan when and where you'll visit.
- Step 10. Begin data collection while remaining flexible to the situation. Adapt methods and adjust questions and activities as warranted.
- Step 11. Record data as collected in a systematic fashion.
- Step 12. Continually analyze data by verifying responses, deepening understanding, and making distinctions and connections between responses.

Case Study

A case study is an in-depth analysis of a particular case – a program, a group of participants, a single individual, or a specific site or location. Case studies can be explanatory, descriptive or exploratory. An explanatory case study can measure causal relationships; a descriptive case study can be used to describe the context in which a program takes place and the program itself, and an exploratory case study can help identify performance measures or pose hypotheses for further evaluation. Case studies rely on multiple sources of information and methods to provide as complete a picture as possible of the particular case.

Advantages of a case study:

- It is good for addressing *how* and *why* questions.
- It is gives concreteness to problems and solutions.
- It is can be used to study evolutionary or decision-making processes.
- It is provides in-depth information on a single setting, group or organization.
- It is can be tailored to specific situations.
- It is can provide background information as a guide for further study.
- It is contributes insight into relationships and personal feelings.
- It is draws out underlying assumptions and general knowledge.
- It is can be used as a supplement to other methods.

Disadvantages of a case study:

- It is time consuming and requires a large amount of data.
- It is may not be generalizable to a larger population
- It is may provide data on only one or two aspects of a problem.
- It is requires good observational, recording and reporting skills.
- Information may be subjective because of investigator bias.

An Example of a Case Study

A detailed and systematic recording of evidence before and after a producer participates in a comprehensive financial farm management program could provide valuable insights into program impact that might be useful in expanding the program to a larger group of producers.

Steps to planning and conducting a Case Study

- Step 1. Review what is expected from the case study.
- Step 2. Define preliminary questions and hypotheses. Hypotheses suggest relationships between variables: e.g., Farmers participate in a program because they derive some benefit from participation.
- Step 3. Identify and define the boundaries of the case.
- Step 4. Assess the ability of the evaluator to carry out a case study: the ability to ask questions, to assimilate large amounts of new information without bias, adaptability and flexibility.
- Step 5. Anticipate key problems, events, attributes and persons that may be encountered.
- Step 6. Form initial plan, including role of on-site observer.
- Step 7. Arrange for access and negotiate plan of action.
- Step 8. Discuss arrangements for maintaining confidentiality.
- Step 9. Make preliminary observations of activities.
- Step 10. Identify informants and sources of information.
- Step 11. Develop record keeping system, files, tapes develop; coding system; arrange for protected storage of data.
- Step 12. Rework priorities based on emerging attributes, problems, events, audiences, etc.
- Step 13. Reconsider issues or other theoretical basis to guide data collection.
- Step 14. Make observations, interview, gather logs, use surveys, etc.
- Step 15. Select vignettes, special testimonies, illustrations.
- Step 16. Classify raw data; begin interpretation of data.
- Step 17. Redefine issues and case boundaries, renegotiate arrangements with hosts, if needed.
- Step 18. Gather additional data, using triangulation to validate key observations, review raw data for various possible interpretations.
- Step 19. Search for patterns in data.
- Step 20. Seek linkages between program arrangements, activities and outcomes.
- Step 21. Draw tentative conclusions, organize according to issues, and organize final report.
- Step 22. Review data, gather new data -- deliberately seek disconfirming evidence.
- Step 23. Describe setting within which case study took place.
- Step 24. Draft report and reproduce material for audience. Consider the report as if it were a story; look for ways the story is incomplete and fill in missing information.

Semi-structured Interviews

Semi-structured interviews with project participants and other key informants begin with an interview guide that lists topics to cover and open-ended questions to ask. Probing techniques are used to solicit answers and raise new topics that reflect the people's perspectives, beliefs, attitudes and concerns.

Advantages of semi-structured interviews:

- They are useful in complex situations in which answers to questions can not be predetermined.
- They can be used to generate hypotheses to guide an evaluation.
- Respondents are not confined by pre-selected choices when answering questions.
- The structure of the interview is not predetermined but develops as the interview unfolds.
- Additional questions can be asked to clarify issues or explore new areas.

Disadvantages of semi-structured interviews:

- They may not be viewed as valid by those more familiar with quantitative methods.
- They require an interviewer who knows when and how to probe for more complete answers and can recognize emerging and relevant topics.

Guidelines for Semi-structured Interviewing

- Identify topics and develop open-ended questions.
- Remember your appearance and mannerisms have an impact on the interview. Dress appropriately and speak in a non-threatening manner and use easy-to-understand terms.
- Select a respondent following the chosen sampling criteria.
- Conduct the interview or select a mutually convenient time to return.
- To avoid distractions, try to conduct the interview without an audience.
- Explain the purpose of the interview to the respondent and remind participants that the interview is voluntary and his/her responses are confidential.
- Establish rapport by beginning with a general conversation on a neutral subject that might interest the respondent and share some personal background and express appreciation for the respondent's responses.
- Begin with simple questions that do not require long answers or a lot of reflection, then move on to more complex and sensitive questions.
- Record answers verbatim.
- Do not express your personal opinions.
- If an answer is incomplete or appears irrelevant, probe to get a clearer response.
- If a respondent refuses to answer a question, attempt to get an answer but avoid doing something that might jeopardize the interview.

Participant Observation

Participant observation entails gathering information about behavioral actions and reactions through direct observation, interviews with key informants, and participation in the activities being evaluated¹. As used in evaluation, the PO evaluator immerses him- or herself in the setting being studied with the intent of understanding the world through the eyes of stakeholders. Participant observation is useful in determining community conflicts or misunderstandings, assessing community needs and problems, and/or identifying means to involve local people in problem solving.

Advantages of participant observation:

- Observation takes place in its natural setting.
- PO is unstructured and flexible.
- It can be readily combined with other methods.
- It's useful for small units such as a neighborhood, a classroom or a group.
- It's useful in assessing long-term effects of programs or practice on local residents.
- It may uncover behavioral patterns, social processes or problematic issues that participants are not aware of.

Disadvantages of participant observation:

- It rarely provides enough information for an evaluation and must usually be combined with other methods such as interviewing.
- It requires an evaluator with well developed observational skills.
- The evaluator has less control over the situation.
- The presence of an evaluator may change the behavior of the group being observed.
- Observations may not be generalizable.
- The observer may lose objectivity as a result of being a participant.
- Time is often a limiting factor.
- It may not be suitable for large and/or heterogeneous groups.

¹ increasingly participant observation emphasizes the observational aspects of the technique and minimizes the participatory aspects.

General Instructions for Engaging in Participant Observation

1. Participant observation (PO) as used in evaluation is motivated by the need to solve practical problems, not to construct theory. Therefore, the evaluator using this technique should enter the field with a initial conceptual framework². The framework should include preliminary issues and the possible relationships between them.
2. Define the main concepts in the framework, e.g. learning style, leadership.
3. Identify sources of information.
4. Select the site in which participant observation is to be carried out. Selecting two or more sites allows for comparative analysis of data. An informal sampling technique is used in PO. The site selected should be representative of the type of program or organization being observed, the organization must be willing to accept the PO evaluator, the PO evaluator must be able to enter into activities under observation, timing is crucial for one time activities, seasonal events, or those with a daily routine.
5. Arrange for access and develop arrangements for maintaining confidentiality.
6. Assemble tools for observation: notebook, pen, camera, tape recorder, etc. How data is recorded depends on the situation. You may want to take notes on the spot or you may want to make notes after finishing your observations. Photographs and tape recorders assist in recording, but in some instances may be intrusive and influence the situation being observed.
7. Begin observation. You do not need to observe everything that is going on, but rather should focus only on those aspects of the activity pertinent to the evaluation. The following questions may help guide your observations:
 - a. What is the setting of the scene you observed?
 - b. What is going on?
 - c. Where are you in relation to the scene you observed?
 - d. What is the content of the situation being observe? – e.g. time of day; weather conditions; approximate number of participants; participants' age, gender, ethnicity, and class; relationships (if any) among participants; people's position in relationships to each other; other activities going on around the scene.
 - e. Are you a participant in the activity? If so, how did your participation affect your observation?
 - f. What is the age, gender, ethnicity, and class of those you are observing?
 - g. If you asked more than one person, did you get the same answer?
 - h. What does the scene you are observing make you think about? What puzzles you? What do you think you understand?
 - i. Is the activity similar to or different from other types of activities you have observed in similar settings?
8. Can the activities that you observe be linked to any theoretical frameworks?

² Casely, David, and Krishna Kumar. 1988. *The Collection, Analysis and Use of Monitoring and Evaluation Data*. John Hopkins Press for The World Bank. Baltimore and London.

Benefit/Cost Analysis

Benefit/cost analysis is typically viewed as an alternative to program evaluation. However, it can also be seen as an extension of the evaluation process. As such, benefit/cost analysis provides a means to systematically quantify and compare program inputs to program outcomes in monetary terms. Valuing both benefits and costs in monetary terms allows them to be directly compared to determine the net impact of the program, make comparisons between alternative programs or projects, assist in program planning, advance organizational accountability and /or expedite program support.

Advantages of a benefit/cost analysis:

- It has high credibility as a source of information.
- It's useful to justify budgets, demonstrate the value of a program, and/or assist in getting the most outcome from program inputs.
- It yields useful information for donors and funders.

Disadvantages of a benefit/cost analysis:

- It may be difficult to quantify costs or benefits in monetary terms.
- It may be difficult to account for opportunity costs, hidden costs and/or assumed costs.
- It may be difficult to account for indirect benefits of the program.
- Bias may occur when assigning monetary value to costs and benefits.
- Bias may occur through underlying and untested assumptions.

Steps to Benefit/Cost Analysis

Step 1. Define the program parameters.

Step 2. Develop a list of costs and benefits from various sources. Program costs include direct costs, implied or indirect costs, and implicit or assumed costs. Program descriptions, professional literature, your own knowledge, and information compiled during initial phases of analysis. Program benefits are the positive outcomes or consequences resulting from the program or project. They include direct benefits and those that accrue over time. When determining costs and benefits, make sure that costs and benefits are measured at the same level.

A. The cost equation: Cost = L + K + I - i

L = labor: The cost per hour for labor including salary and fringe benefits. Fringe benefits vary but normally fall within 22 to 35 percent of full salary. The full labor hourly formula (L) is: $(S+S.35)/260/8$ where S = salary and S.35 = fringes, 260 = workdays per year and 8 = hours per workday.

K = direct costs: Direct program costs budgeted for or assigned to the program, e.g., supplies, correspondence, communications, travel and per diem expenses, equipment and audiovisuals. If costs are shared between projects, the total is calculated from a cost/share equation. Opportunity costs are defined as opportunities that participants have lost in order to participate in the program. Opportunity costs are included in direct costs to the participants, the presenters or the stakeholders, depending on the level of analysis.

I = indirect costs: Costs indirectly associated with the participants but directly associated with the program e.g., administrative costs such as facility rental, photocopying, report costs, telephone and prorated equipment and supplies costs.

i = discount amortization: Measurable returns over time (both positive and negative). Discount amortization is not included if returns can not be traced over time.

B. The benefit equation: B = Cr + DB + IB

Cr = cost reductions attributable to program activities.

DB = direct benefits - the primary outcomes experienced by participants and others directly involved in the program. They are typically derived from program objectives.

IB = indirect benefits - secondary or intangible outcomes of the program or project experienced by participants, non-participants or society in general. These outcomes or consequences can be positive or negative.

Step 3. Compare costs with benefits, either directly by subtracting costs from benefits or as a ratio of benefit to cost. The first equation provides a means of comparing costs with benefits within a given program; the second allows comparison between programs.

Sample Sheet for Benefit/Cost Analysis

Benefits Estimate Worksheet

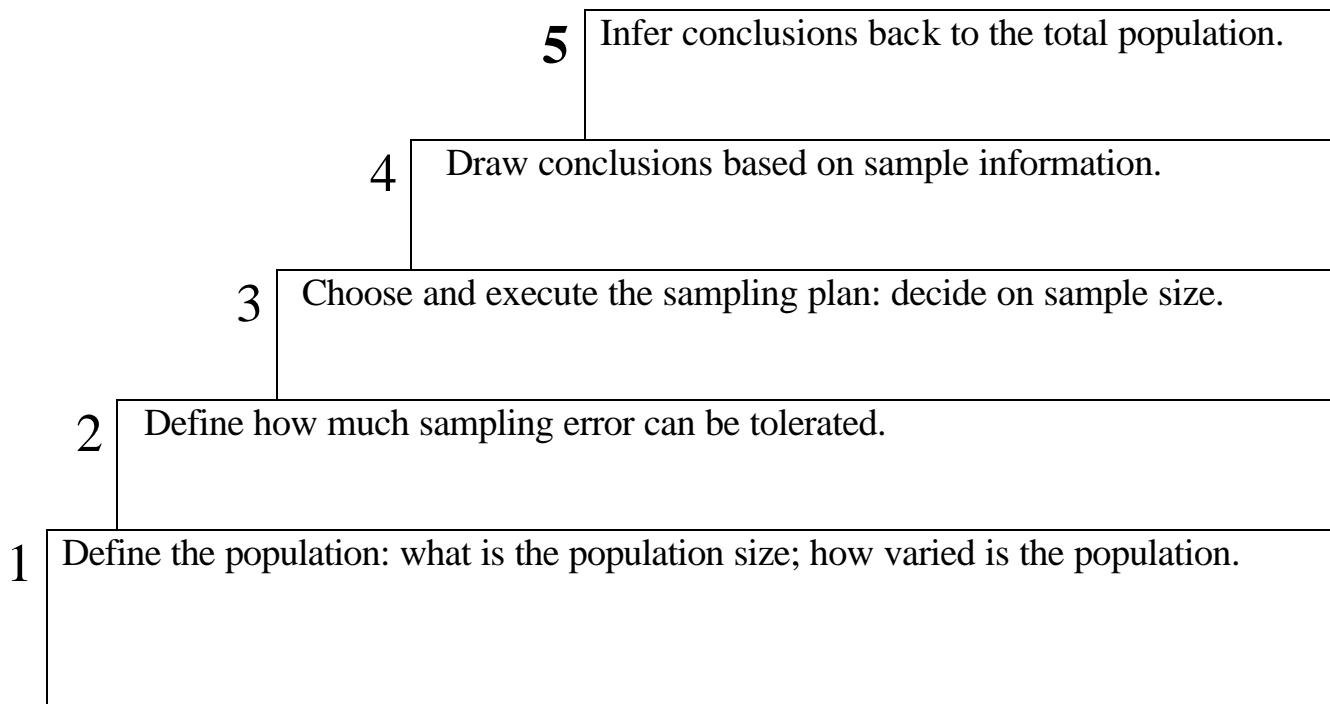
Cost Estimate Worksheet

	Estimated Benefits:				Resources needed in Units	Estimated cost:		
Number of beneficiaries						No. of units	Unit value	Total cost
Direct benefits					Direct costs			
1.					Labor	Hours:		
2.					1.			
3.					2.			
4.					3.			
5.					Direct costs			
6.					1. Rent			
					2. Utilities			
Indirect benefits					Equipment & materials			
1.					1. Printed materials	Pieces:		
2.					2. Furnishings			
3.					3. Instructional Materials			
4.					4. Travel	Miles:		
5.					Opportunity costs			
6.					1. Child care			
					2. Food			
					3. Travel			
					Indirect costs			
					Total program costs			
Total program benefits								
Benefit/cost ratio								

Step 7. Sampling for Evaluation

A sample is a set of respondents selected from a larger population for the purpose of a survey. When done properly, the sample represents the characteristics of the population as a whole. Sampling saves time, money, materials and efforts without sacrificing accuracy and precision.

Five Steps in Sampling



Sample Size

How large should a sample be? A sample size of 100 respondents is often cited as a minimal number for a large population. The practical maximum size is about 1000 respondents. Generally, a sample of fewer than 30 respondents will not provide enough certainty to prove useful. However, several factors need to be considered when determining actual sample size.

Characteristics of population – addresses the amount of variability in the population to be sampled. A relatively homogeneous population may permit a smaller sample size. Conversely, a more heterogeneous one may require a larger population size.

Sampling error - the difference between an estimate taken from the population and that taken from the sample when the same method is used to gather the data. Sampling error is larger when the sample size is small. It is therefore advisable to use the largest sample size possible given the constraints on time, money and materials.

Degree of precision - measures the degree to which an estimate approximates the estimate obtained from the total population, assuming the same method of data collection was used. In designing a sample, the evaluator may begin by defining the degree of precision desired.

Margin of error - It is a matter of choice depending on the objectives of the inquiry. If we want to be relatively safe about our conclusions, then a 5 percent margin of error is acceptable (see appendix). In general, more subjects are needed for a .01 alpha test than a .05 alpha test, and a two-tailed test requires a larger sample size than a one-tailed.

Confidence level - the probability that a value in the population is within a specific, numeric range when compared with the corresponding value computed for the sample. Generally, a 95 percent confidence level will give the security needed to draw conclusions for the larger population based on the sample.

Cost - A small sample size reduces cost.

Table for Determining Sample Size from a Given Population

n*	s*	n	s	n	s
10	10	220	139	1200	291
15	14	230	143	1300	296
20	19	240	147	1400	301
25	24	250	151	1500	305
30	28	260	155	1600	309
35	32	270	158	1700	313
40	36	280	161	1800	316
45	49	290	165	1900	319
50	44	300	168	2000	322
55	48	320	174	2200	327
60	51	340	180	2400	331
65	55	360	185	2600	334
70	59	380	191	2800	337
75	62	400	195	3000	340
80	66	420	200	3500	346
85	69	440	205	4000	350
90	72	460	209	4500	353
95	76	480	213	5000	356
100	79	500	217	6000	361
110	79	550	226	7000	364
120	91	600	234	8000	366
130	97	650	241	9000	368
140	102	700	248	10000	369
150	107	750	254	15000	375
160	112	800	259	20000	377
170	117	850	264	30000	379
180	123	900	273	40000	380
190	127	950	277	50000	381
200	131	1000	284	75000	382
210	135	1100		1000000	384

* n = population size; s = sample size

Sampling Techniques

Random or probability sampling is based on random selection of units from the identified population. Random sampling techniques include:

Simple random sample - all the individuals in the population have an equal and independent chance of being selected as a member of the sample. A random numbers table is sometimes used with a randomly selected starting point to identify numbered subjects (see appendix).

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.

Stratified sampling – is used to ensure that certain subgroups in the population will be represented in the sample in proportion to their numbers in the population. Each subgroup is separately numbered and random selection is used for each subgroup. A definite rationale should exist for selecting any such subgroup.

Cluster sample - the unit of sampling is not the individual but rather a naturally occurring group of individuals such as a classroom, organization or community.

Matrix sample - one sample of people receives a given sampling of questions and another sample of people receives another sampling of questions.

Purposive sample is chosen to include a wide variety of people on the basis of a number of specifically chosen and critical characteristics. Purposive sampling does not rely on random selection of units.

Accidental sample - sample consists of individuals who are available at the time. This is the weakest type of sample. Generalizations to the larger population can not be made.

Reputational sample people are selected to respond to a survey or an interview based on a judgment of who is and who is not a “typical” representative of the population.

Step 8. Collecting, Analyzing & Interpreting Data

Various kinds of data analysis exist for both quantitative and qualitative data. You should consider whether the analyses would provide the information needed to answer the questions posed by the evaluation and the analytical skills the evaluator possesses.

Qualitative Data Analysis

Analysis and interpretation of qualitative data are not simple technical processes like the analysis of quantitative data. Analysis of qualitative data is the process of bringing order to the data and organizing what there is into patterns, categories and basic descriptive units. Interpreting qualitative data is the process of bringing meaning to the analysis, explaining patterns, and looking for relationships and linkages among descriptive dimensions. The evaluator and/or stakeholders then make judgments about assigning value or worth to what has been analyzed and interpreted³.

Characteristics of qualitative data analysis:

- It begins as soon as data collection begins.
- It is an iterative process that continues throughout data collection.
- Issues of validity and reliability are expressed in terms of clarity, verifiability and replicability.

When doing qualitative analysis consider:

- The words used by the participants and the meaning of those words.
- The context. Interpret the comments in light of the context.
- The internal consistencies and inconsistencies. Determine the cause of inconsistencies.
- The frequency or extensiveness of comments.
- The intensity of comments.
- The specificity of response.
- Dominant themes.

³ EDGE Guide to Evaluation, Ohio Cooperative Extension Service, Ohio State University.

Types of Qualitative Data Analysis

Content Analysis: A coding or classifying technique that investigates pattern of information and the meaning of data within a specific conceptual framework.

Content analysis

Content analysis is a research method that uses a set of procedures to make valid inferences from text such as newsletters, meeting minutes, correspondence, interview transcripts, etc. The inferences are about the sender(s) of the message, the message itself, or the audience of the message. Content analysis can be used for many purposes, such as auditing communication content against objectives, coding open-ended questions in surveys, describing attitudinal and behavioral responses to communications, revealing the focus of individual, group, institutional or societal attention toward something. A central idea in content analysis is that many words of text are classified into much fewer content categories. Each category may consist of one, several or many words. Words, phrases or other units of text classified in the same category are presumed to have similar meanings. Content analysis procedures create quantitative indicators that assess the degree of attention or concern devoted to cultural units such as themes, categories or issues. The investigator then interprets and explains the results using relevant theories. It involves three steps:

1. Measurement consists of counting the occurrences of meaning units such as words, phrases, content categories or themes. Counting generates results that allow for more precise comparisons among texts. We also want to know how much more (or less) attention is devoted to some issue than to others. Quantitative analytical procedures often reveal similarities and differences among texts that would be difficult, if not impossible, to detect otherwise.
2. Representation deals with the fact that in content analysis essential syntactic or semantic features of the languages or text are omitted, as it is difficult to encode or represent the richness of language. One way that the meaning of words, phrases, or other textual units is represented is through classification into a set of categories. In assigning meaning to categories, not all connotations or nuances of meaning are pertinent. Examples: kind, state.
3. Interpretation

Reliability

To make valid inferences from the text, it is important that the classification procedure be reliable in the sense of being consistent: Different people should code the same text in the same way. Reliability problems in content analysis usually grow out of the ambiguity of word meanings, category definitions, or coding rules. Three types of reliability are pertinent to content analytic analysis: stability, reproducibility, and accuracy.

Stability refers to the extent to which the results of content classification are invariant over time, i.e., whether content will be coded in the same way if it is coded more than once by *the same coder*.

Reproducibility refers to the extent to which content classification produces the same results when the same text is coded by *more than one coder*.

Accuracy refers to the extent to which the classification of text corresponds to a standard or norm. It is the strongest form of reliability, but usually not available and done. Sometimes, it is used to train coders, though.

Validity

The classification procedure must also generate valid variables, that is, it must measure or represents what the investigator intends it to measure. As happens with reliability, validity problems also grow out of the ambiguity of word meaning and category or variable definitions.

Face validity (weakest) consists of the correspondence between the investigators' definitions of concepts and their definitions of the categories that measured them. A category has face validity to the extent that it appears to measure the construct it is intended to measure.

A measure has construct validity to the extent that it is correlated with another measure of the same construct. Thus, construct validity entails the generalizability of the construct across measures or methods. There is no simple right way to do content analysis, investigators must judge what methods are most appropriate for their purpose. Large portions of text, such as paragraphs and complete texts, usually are more difficult to code as a unit than smaller portions, such as words or phrases, because large units typically contain more information and a greater diversity of topics. Hence, they are more likely to present coders with conflicting cues.

Creating and testing a coding scheme

1. Define the coding units: Words, word sense (code different senses of words with multiple meanings or code phrases that constitute a semantic unit), sentences (when interested in words or phrases that occur closely together), themes, paragraphs, whole text.
2. Define the categories, which involves two decisions: First, whether the categories are mutually exclusive, and second, how narrow or broad the categories are to be.
3. Test coding on a sample of text.
4. Assess reliability.
5. Revise the coding rules.
6. Return to step three until coders achieve sufficient reliability.
7. Code all the text.
8. Assess achieved reliability. Coders are subject to fatigue and are likely to make more mistakes as the coding proceeds. Also, as the text is coded, their understanding of the coding rules may change in subtle ways that lead to greater unreliability.

INTERPRETING DATA ANALYSIS

Data analysis focuses on organizing and reducing information and making logical or statistical inferences; interpretation, on the other hand, attaches meaning to organized information and draws conclusions. All interpretations, to some extent, are personal and idiosyncratic. Therefore, not only interpretations but also the reasons behind should be made explicit. Useful interpretation methods include the following:

1. Determining whether objectives have been achieved.
2. Determining whether laws, democratic ideals, regulations, or ethical principles have been violated.
3. Determining whether assessed needs have been reduced.
4. Determining the value of accomplishments.
5. Asking critical reference groups to review the data and to provide their judgments of successes and failures, strengths and weaknesses.
6. Comparing results with those reported by similar entities or endeavors.
7. Comparing assessed performance levels on critical variables to expectations of performance or standards.
8. Interpreting results in light of evaluation procedures that generated them.

One method of bringing multiple perspectives to the interpretation task is to use *stakeholder meetings*. Stakeholders can be supplied in advance with the results, along with other pertinent information such as the evaluation plan and list of questions, criteria, and standards that guided the evaluation; that way, the meeting can be devoted to discussion rather than presentation. At the meeting, findings are systematically reviewed in their entirety, with each participant interpreting each finding, using questions such as: What does this mean? Is it good, bad or neutral? What are the implications? What, if anything, should be done?

Quantitative Data Analysis

Simple statistical analysis

Scales of measurement:

Scales of measurement refers to the type of variable being measured and the way it is measured. Different statistics are appropriate for different scales of measurement. Scales of measurement include:

Nominal: mutually exclusive and logically exhaustive categories.

Examples: marital status; gender; group membership; religious affiliation.

Ordinal: ranked or ordered.

Examples: letter grades; social class; attitudinal variables.

Interval: ranked and ordered in standard units of measurement.

Examples: years of age; degree; calendar year; scores on a test; IQ.

Ratio: an interval scale with an absolute zero starting point.

Examples: years of age; years of education; time; length; weight.

Analyzing descriptive data:

Measure of central tendency:

The purpose of central tendency is to report a single summary score or category that best describes a set of observations. Mean, median and mode are the most common measurements of central tendency and are used to compare one group with another, identify some behavior that is unknown, or compare a group to a standard.

The mean is used for interval variables. It is the arithmetic average of all observations. You calculate mean by totaling all observations (scores or responses) and dividing by the number of observations. The mean is sensitive to “outliers” or extreme values in the observations. When your data has a few extremely small or large observations, the data are “skewed.”

Example: 15 participants received the following scores: -2, -1, 1, 4, 4, 4, 7, 7, 7, 7, 7, 8, 8, 8, 9.

The mean of the scores ($\Sigma X/n$) is:

$$(-2)+(-1)+(1)+(4)+(4)+(4)+(7)+(7)+(7)+(7)+(7)+(8)+(8)+(8)+(9)/15 = 5.1$$

The median is most appropriate for ordinal variables. The median is the middle observation. Half of the observations are larger and half are smaller. The median is not as sensitive to the outliers as the mean.

Examples: Observation 1 = 6,8,13,18,25. The median is 13, because half the scores fall above this number and half fall below; Observation 2 = 1, 4, 7, 8, 10, 11, 21, 22. The median is determined by summing the middle two numbers, 8 and 10, and dividing by 2. The median is 9.

The mode is used for nominal variables. It is the observation or category that occurs most frequently. The mode can be used to show the most “popular” observation or value. A distribution can be either unimodal or bimodal.

Distribution A		Distribution B	
Score	Frequency	Score	Frequency
23	2	33	1
45	6	21	7
34	8	61	21
25	11	75	4
73	15	66	3
83	18	24	7
54	10	74	10
66	12	88	21

Distribution A is unimodal or has a single mode of 83, with 18 responses.

Distribution B is bimodal or has two modes, 61 and 88, with 21 responses each.

When to Use Mean, Median or Mode

Use the mean when:

- The distribution is approximately symmetrical.
- You are interested in numerical values.

Use the median when:

- You are concerned with the typical score.
- The distribution is skewed.
- You have ordinal data.

Use the mode when:

- The distribution has two or more peaks.
- You want the prevailing view, characteristic or dominant quality.

Test for Differences

Chi-Square (χ^2) is the most popular of all non-parametric inferential statistical methods. Chi-square tests for differences between categorical variables (e.g., nominal or ordinal data). There are both “one-way” and “two-way” chi-square procedures.

Example of one way chi-square: A sample group is asked a question about political party preference, assuming the question on the instrument form requires a categorical response (Democratic, Republican, Independent, etc.). The one-way chi-square would test for differences in popularity between the political party categories relative to the sample’s response to the question.

Example of two-way chi-square: Used if two categorical variables are to be compared. If the same group above were split into male and female, thus creating a new variable, “sex of the respondent,” then this categorical variable could be compared (or “cross-tabulated”) to political party choice. In this way, comparisons between the sexes on political preference may be evaluated (e.g., significantly more males are Republican and more females are Democratic).

Both the one-way and two-way chi-square procedures result in a chi-square value and associated significance (probability) level. Chi-square is a non-parametric statistic and as such requires no parametric data assumptions. The data must be categorical in nature.

t-Test is used to test the difference between two means even when the sample sizes are small. The significance of the t statistics depends upon the hypothesis the researcher plans to test. If you are interested in determining whether there is a significant difference between two means, but you do not know which of the means is greater, use the two-tailed test. If you are interested in testing the specific hypothesis that one mean is greater than the other, use the one-tailed test. Data should satisfy parametric assumptions: 1) the sample is selected from populations that are nominally distributed; 2) there is homogeneity of variance -- i.e., the spread of the dependent variable within the group tested must be statistically equal; and 3) data are of continuous form with equal intervals of quantity measurement. Dependent variables must be interval or ratio-type data.

T-test for matched pairs: if both groups of data are contained in each data record, the appropriate t-test is for matched pairs. An example of an appropriate use of the t-test for matched pairs might be to compare pretest and posttest scores where each person took a pretest (variable 1) and a posttest (variable 2). Both values are contained in each data record.

T-test for independent groups: If each case in the data file is to be assigned to one group or the other based on another variable, use the t-test for independent groups. For example, to compare reading scores between males and females, split the reading scores into two groups, depending on whether the person is male or female (each record in the date file is assigned to one group or the other).

Degrees of freedom: (*this is not a complete description*) The degrees of freedom (d.f.) reflect sample size. When two independent samples are being considered, d.f. are equal to the sum of two sample sizes minus 2; i.e., $d.f. = (n_1 + n_2 - 2)$.

Measures of variance indicate the spread or dispersion of the group and include range, variance and standard deviation.

Range is the difference between the largest and the smallest scores in a distribution.

Example: Scores of 3, 6, 8, 10, 14, 17. The range is 14 points. The scores range from 3 to 17.

Variance is the mean of the squares of the deviation scores. Calculate the difference (deviation) between each score and the mean of the scores, square the deviations, sum the squares and divide the sum by the number of scores minus 1.

Standard deviation measures the spread of data about their mean and is an essential part of any statistical test. It is calculated by taking the square root of the variance. This transforms variance into the same unit of measurement as the raw scores. Standard deviation is expressed in terms of “one standard deviation above the mean” or the like. If the standard deviation is 11 and the score is 63, then one standard deviation is above the mean is 74, two standard deviations is 85 and so forth. The value of this figure becomes apparent when we understand the relationship between standard deviations and percentiles in a normal curve. The area contained within +1 and - 1 standard deviations of the mean includes approximately 69 percent of all scores on the distribution. Therefore, in our earlier example 68 percent of all scores were between 52 and 74.

Another way of assessing the meaning of the standard deviation is to compare scores with percentiles. It is known that, in a normal distribution, 97.7 percent of the cases are below two standard deviations above the mean. So when a raw score for one case is found to be two standard deviations above the mean, we know that the case scored higher than 97.7 percent of all other cases.

Selection Guide for Common Statistical Methods

Data Type	Statistical Method	Differences (between groups)	Testing for: Relationships (within one group)
CATEGORICAL	Nominal		Contingency Coefficient
	Non-parametric	Chi-square	
CONTINUOS	Ordinal		
	Interval	ANOVA (3groups)	Pearson Correlation Multiple Regression Discriminate Analysis*
	Parametric		
	Ratio	T-test (2 groups)	

*If variable to be predicted is categorical

Step 9. Communicate Findings

Evaluators have a responsibility to report their findings to stakeholders and other audiences who may have an interest in the results. Communication with stakeholders should occur throughout the evaluation process to help ensure meaningful, acceptable and useful results. (Continue this)

Reporting plan: Developing a reporting plan with stakeholders can help clarify how, when and to whom findings should be disseminated.

- *Who are the intended audiences?*
- *What information will be needed?*
- *When will information be needed?*
- *What reporting format is preferred?*

Reporting results: A variety of reporting procedures may be used.

- ◆ *Verbal reports.*
- ◆ *Short communications.*
- ◆ *Executive summary.*
- ◆ *Public meetings.*
- ◆ *Personal discussions.*
- ◆ *Audio-visuals.*
- ◆ *Journal or newspaper articles.*
- ◆ *Graphs, tables and charts.*
- ◆ *Newsletters, bulletins, and brochures*
- ◆ *Poster sessions.*
- ◆ *Question-answer periods.*

An Evaluation report usually contains:

- ◆ *A description of :*
 - *The program.*
 - *The program setting.*
 - *The purpose of the evaluation.*
 - *The procedures used.*
- ◆ *A summary and analysis of findings.*
- ◆ *An explicit justification of conclusions.*
- ◆ *Recommendations for future changes.*

Reporting Tips

- Reports that are short, concise and to the point are the ones that get attention.
- Craft the style and content of the evaluation report to fit the intended audience.
- Avoid technical terms that your audience may not know.
- Use a conversational tone.
- Use a combination of long and short sentences.
- Read report aloud to check for confusing ideas and sentences.
- Write in an active voice.
- Use a logical structure for your documents.
- Allow sufficient time for writing drafts and getting feedback and proofreading.

Reporting Negative Findings

At times you may be called on to report negative findings - the program may not have met its objectives, the program is being mismanaged or changes are needed. Evaluation can both identify and point to the causes of negative results.

Reporting these difficulties can help avoid future mistakes and suggest ways to improve. However, negative findings must be reported in a manner that helps promote 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,” “things to work on”.
- Creates an atmosphere of reflection, dialogue and positive thinking.
- Helps stakeholders think of themselves as problem solvers.
- Communicates with stakeholders throughout the evaluation process.
- Helps stakeholders process negative findings.

Step 10. Applying and Using Findings

An evaluation should not be considered complete until the findings of the evaluation are applied:

To make decisions about program continuation.

To improve on-going programs,

To plan future programs.

To inform program stakeholders.

When evaluators are evaluating their own programs there are fewer problems involved in implementing findings, however, where evaluators are not the persons conducting the program, the likelihood of evaluation findings being ignored is greater. When the concerns of stakeholders have been incorporated into the evaluation process, evaluation findings are more likely to be used.

A Final Step: Evaluating the Evaluation

Evaluating evaluation differs little from the actual process of the evaluation itself. It must meet the same standards and follow similar steps as the original evaluation. Evaluating evaluation considers:

- **Conceptual clarity** - Is the evaluation well focused and the purpose, role, and general approach clearly stated?
- **Description of object to be evaluated** - Does the evaluation contain a thorough, detailed description of what is evaluated?
- **Recognition and representation of legitimate audiences** - Have all legitimate evaluation audiences had a voice in focusing the study and an opportunity to review results?
- **Sensitivity to political problems in evaluation** - Has the evaluation been sensitive to and coped satisfactorily with potentially disruptive political, interpersonal, and ethical issues?
- **Specification of information needs and sources** - Did the evaluation specify needed information and sources of information?
- **Comprehensives & inclusiveness of data** - Has the evaluation collected date on all important variables and issues, without getting bogged down in inconsequential data?
- **Technical adequacy** - Did the evaluation design and procedures yield information that meets scientific criteria of validity, reliability, and objectivity?
- **Appropriate methods and analysis** - Were the appropriate methods chosen for the evaluation? Were they used correctly? Were data analyzed and interpreted carefully?
- **Consideration of costs** - Did the evaluation adequately consider cost factors along with other variables?
- **Explicit standards & criteria for judging the evaluation** - Did evaluation contain an explicit listing and/or discussion of the criteria and standards used to make judgments about the evaluation object?
- **Judgments and/or recommendations made by evaluation** - Did the evaluation go beyond just reporting findings but also offer judgments and recommendations suggested by the data? (add these to document)
- **Reports tailored to audiences** - Were the evaluation reports provided at appropriate times and in appropriate formats to the appropriate audiences?

Evaluating Evaluation: Hierarchy of Evaluation Accountability

Program Chain of Events	Matching Level of Evidence
Program & decision impacts	To what extent and in what ways could the program improved? To what extent were informed, high-quality decisions made?
Practice and program changes	To what extent did intended use occur? Were recommendations implemented?
Stakeholder's knowledge and attitude changes	What did intended users learn? How were users' attitudes and ideas affected?
Reactions of primary users	What do intended users think about the evaluation? What's the evaluation's credibility? believability? relevance? accuracy? potential utility?
Stakeholder participation	Who was involved? To what extent were key stakeholders and primary decision makers involved throughout?
Evaluation activities	What data were gathered? What were the focus, the design and the analysis? What happened in the evaluation?
Inputs	To what extent were resources for the evaluation sufficient and well managed? Was there sufficient time to carry out evaluation?

Evaluation Planning Worksheet

Identify the program to be evaluated, its objectives and stakeholders	Assess the feasibility of implementing an evaluation	Consult Stakeholders to clarify indicators of program merit	Identify approaches to data collection	Select data collection techniques

Identify target population and select sample	Who will collect data?	How will data be analyzed and interpreted?	How will evaluation findings be shared with stakeholders?

References

- Archer, T. and Layman, J. (1991). "Focus Group Interview" Evaluation Guide Sheet, Ohio Cooperative Extension Service.
- Bennett, Claude F. and S. Kay Rockwell (1995). *Targeting Outcomes of Programs (TOP): An Integrated Approach to Planning and Evaluation*. Lincoln, NE: Cooperative Extension, University of Nebraska.
- Bennett, Claude. 1979. *Analyzing Impacts of Extension Programs*. Washington, DC: US Department of Agriculture, Science and Education Administration (ES C-575).
- Case R.; Andrews, M. and Werner, W. (1988). How can we do it? an evaluation training package for development educators. British Columbia, Canada: Research and Development in Global Studies.
- Contant, C. K. (1993). "Assessing What and Why: Designing and Using Evaluations Effectively for Local Level Programs." Paper presented at the Rural Nonpoint Source Pollution in the Upper Midwest Conference, March.
- 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. Thousand Oaks, California: Sage.
- Fraenkel J. R and Wallen, N. E. (1996). How to Design and Evaluate Research in Education. New York: McGraw-Hill Inc.
- Krueger, R.A. (1994). Focus Groups: a Practical Guide for Applied Research. 2nd edition, Thousand Oaks, California: Sage.
- Mueller, D.J.(1986). Measuring social attitudes. New York: Teachers College Press.
- Neito, R. and Henderson, J.L. (1995). Establishing Validity and Reliability (draft). Ohio State Cooperative Extension.
- Patton, M. Q. (1997). Utilization-focussed evaluation: The New Century Text. Newbury Park, California: Sage.
- Salant, P. and Dillman, D.A. (1994). How to conduct your own survey. New York, NY: John Willey & Sons, Inc.
- Wholey, J. S.; Harty, H. P. and Newcomer, K. E. (eds.). (1994). Handbook of practical program evaluation. San Francisco: Jossey-Bass Publishers.
- Worthen, B. R. and Sanders, J. R. (1987). Educational evaluation: alternative approaches and practical guidelines. New York: Longman.
- Yin, R. K. (1984). Case study research: design and methods. Applied Social Research Methods Series. Vol. 5. Newbury Park, California: Sage.